

## More focused agenda for Unesco

**Britain will probably not rejoin Unesco at this stage, but has a right (and responsibility) to press for new terms of reference as the price for its subscription.**

BRITAIN and the United States had good reason for walking out of Unesco nearly a decade ago. The organization had fallen into spectacular administrative chaos, its work programme was disjunct to say the least, and its senior management was perversely determined on a project to launch a 'new world order' in communications that had all the appearances of being midway between an international system for the management of news and a recipe for its control and censorship. Now, by general consent, Unesco is in better shape. Dr Federico Mayor, the present director-general, has improved the administration and made the work programme more pointed. So should not Britain and the United States now rejoin as a mark of their good faith? The question, which Unesco has been asking, now confronts the British government. The United States, when asked, is said to have replied that it was confident that the US Congress would not vote the necessary funds, and has been let off the hook for the time being.

Not surprisingly, the question follows by only weeks the celebrations of the fiftieth anniversary of the United Nations as a whole. If there is a good time to ask lapsed members to rejoin, this must be it. There are general difficulties. First, the Republican Congress in the United States has declared an interest in an audit of the other UN specialized agencies (which is why the administration is confident that it could not get the funds to rejoin Unesco). Second, rumblings of discontent about the United Nations as a whole were all too loud and clear in the run-up to the anniversary. The role and membership of the Security Council will be on the diplomatic agenda for some time to come, various members will be sucking their teeth about the UN's peacekeeping operations for several months and the efficiency (and the cost) of the organization as a whole will be an issue for even longer. Unesco membership is a separate issue, but is bound to be confused with the others.

### Contradictions

Unesco's particular problem is that the constitutional contradictions that led to the rows of the 1980s still persist. Like the other specialized agencies, Unesco has a broad charter. Its goals are to foster the three elements named in its title — science, education and culture. But what it does in practice is determined by the voting members, who are predominantly developing countries. That means that work programmes (like those of other specialized agencies) are skewed to the interests of the developing world. In principle,

there is nothing wrong with that; in many contexts, the need for development assistance should take precedence over most other good works. The result is to turn most of the specialized agencies into providers of development assistance — much of Unesco's scientific work is an attempt to draw universities in developing countries into collaboration with better-established institutions. It is useful work, but it diverts Unesco from the projects it is uniquely equipped to handle.

### Projects

What are they? There are many practical issues in which Unesco has (and should) play a leading role. Since the Second World War, it has been midwife to the Berne Copyright Conventions, which regulate the use of printed copyrights more or less worldwide. In some respects (but not all), that is now a rich countries' problem, but eventually poor countries will also have a stake. (In the business of intellectual property more generally, Unesco had a natural role, but ceded that to various bilateral negotiators and the World Trade Organization.) Over the years, Unesco has been the chief source of support for the International Council of Scientific Unions, in no sense a prisoner of the developing world. It has also usefully been designating heritage sites, which fortunately are distributed worldwide. Nobody should forget that Unesco was one of the prime movers in launching the International Geophysical Year at the end of the 1950s.

In none of these activities can it be said that developing countries have skewed the pattern of work. The trouble is that there are many others that lie in limbo. Patenting policies (especially in genetics) need authoritative and continuing scrutiny. Unesco has said little about information problems since the Internet became everybody's plaything. What is to be done about the restrictions on the free movement of scholars (not only scientists) when they have legitimate investigations to carry out? On the principle that the scholarly community is truly international, what can be done to safeguard crucial documents and other evidence? Is scientific misconduct an international problem, and what should be done about it? There is a host of matters like these that need attention. (Mayor has set up a potentially useful Biomedical Ethics Commission.) In general, Unesco could be the international academic community's champion in all these matters. It is not.

That is not Unesco's fault. For one thing, the developed



countries' votes would not favour such a concentration on general problems. For another, Unesco is probably ill-equipped to do the work. But there is a case for saying that the organization's terms of reference (and those of the other agencies as well) should be made more specific, and stated in such terms that the important issues cannot be neglected. If the British government (as seems likely) says "No, not yet!" to Mayor, his near namesake, Mr John Major, might raise with him the question of more specific terms of reference. Mayor may say that Major is merely moving the goalposts. Major can fairly reply that some steps of this kind are the way forward for all the agencies of the United Nations. □

## Go fish for patents

**An academic group has protested against gene patenting by putting a nucleotide sequence on the Internet.**

FROM time to time, people throw spanners in the conventional works with good effect. That is what the collaboration hunting for the presumed second breast-cancer gene have now done (see page 425). Those concerned, from the Washington University, St Louis, and the Sanger Centre at Cambridge, England, have published a large stretch of the nucleotide sequence of human chromosome 13, allowing the world at large to hunt for the presumed tumour suppressor gene (mutations of which are supposed to lead to breast cancer). The lucky winner will be able to claim the patent that goes with a definition of the gene and an account of the mutations therein linked with actual cases. "What philanthropy!", it may be said.

Or is it? This little stratagem vividly illustrates the frustration that has built up among researchers at the way prudent concern for patent rights impedes serious work. Admittedly, it will not be child's play to find the gene concerned in the 900,000 base-pair sequence now made public, but that is a lot better than simply knowing that it is somewhere on chromosome 13. Yet there are many laboratories in which all else would have been abandoned for a search for this potentially valuable (or at least money-spinning) gene. What the Cambridge-St Louis team has said is that they, personally, cannot be bothered.

The group has also made a general point of some importance. Having defined the sequence that incorporates the gene, and on the assumption that its eventual discovery will be of some medical benefit, the group is seeking to make the point that sequence data as such is 'pre-competitive' and that it should be quickly made available. It will be interesting to see how many other groups follow suit. If there are none, or only a few, that will give molecular genetics a bad name, for there is no question that delays of this kind will impede therapeutic benefits, not to mention the equally important goal of knowing the whole sequence of the human genome. Unless, that is, the group is really looking for the gene and has distributed 900,000 base pairs of irrelevant sequence in the hoping of throwing others off the scent. □

## So God plays dice!

**Professor Stephen Hawking's announcement that the Universe is unpredictable is a little premature.**

PROFESSOR Stephen Hawking, who holds what was Isaac Newton's chair at the University of Cambridge, filled the Albert Hall in London last week. That is an achievement in itself; the event was a lecture on physics. The box-office money will go to a charity for those with motor-neuron (or Lou Gehrig's) disease, from which Hawking has suffered for more than twenty years. He is also the author of the now-famous best-seller, *A Brief History of Time*. Personally, he is a brave man. Although confined to a wheelchair and able to speak only with a voice-synthesizer, he travels indefatigably in pursuit of his academic interests and as a public speaker.

Hawking was talking to a general audience on the theme that people have come to identify with him — the Universe and all that. He argued against the Cartesian view that the Universe is a machine whose behaviour is essentially predictable. It is not just that there is quantum mechanics and the Uncertainty Principle. More recently, there has emerged from post-Newtonian limbo the concept of deterministic chaos, already a headache for those who would predict the movement of objects in the Solar System. But Hawking has his own particular contribution to the alleged unpredictability of the Universe — black holes.

The argument goes back to Hawking's own work, which showed that the vacuum in the neighbourhood of a black hole would be polarized and would become the source of pairs of fundamental particles, electrons and quarks (see *Nature* **248**, 30–31; 1974). Because some of these material particles would escape the gravitational field that created them, there would be a net loss of material from the system, the black hole would steadily lose mass by means of radiation, more quickly as its mass decreased, until it eventually evaporated. Some have speculated that this process may account for the puzzling  $\gamma$ -ray bursts seen sporadically in the sky.

There are other odd features of this process. Because the matter that enters a black hole is more structured than that which eventually escapes, black holes are devices for getting rid of information and thus of creating entropy, presumably at the expense of the entropy of other parts of the Universe. This, Hawking said last week, is where the unpredictability of the Universe comes from: Einstein's exasperated observation that "God does not play dice!" does not apply.

The message will have comforted many of those present. Descartes and Laplace never struck a chord in Britain. But it is also a misleading message, and will remain so while the Hawkings of this world have not succeeded in their serious quest for a unification of quantum theory and gravitation. It seems inevitable that there are massive objects which, by virtue of their mass, are more dense even than neutron stars, but whether they are the literal mathematical singularities must still be an open question. Whether a general audience will have been helped by Hawking's account is another. □

# Open access to sequence data 'will boost hunt for breast cancer gene'

**London.** A sequence of 900,000 nucleotide bases, thought to contain a gene involved in susceptibility to breast and perhaps other cancers, was made publicly available on the Internet last week by scientists at Washington University, St Louis, Missouri, and the Sanger Centre in the United Kingdom.

The move is intended to help research teams on both sides of the Atlantic to locate and identify the gene in question, BRCA2 (breast cancer 2). "The availability of this sequence data will be a significant boost to the search for the gene," says Mike Stratton of the UK, Institute of Cancer Research (ICR), one of the scientists who last year announced that the gene appeared to lie somewhere on chromosome 13.

At the same time, the decision to make the sequence openly available, with no restrictions on how the information is used, reflects the view of those supporting the sequencing work — which include the Cancer Research Campaign and the Wellcome Trust in the UK, as well as the US National Institutes of Health and Department of Energy — that such information should not be treated as private property.

"I believe that the recent scramble to try to lock up [the patent rights to] the sequence of a gene to a single end, and with a single company, is profoundly counterproductive," says John Sulston, director of the Sanger Centre, in a thinly veiled reference to last year's controversy over the decision by the company Myriad Genetics and the University of Utah to apply for a broad patent on the discovery of another breast cancer gene, BRCA1 (see *Nature* 372, 118; 1994).

"We think the best way to help people to search for and understand the function of the gene is by releasing widely and freely information about the sequence in which it appears to be contained," says Sulston. "The sequence should be openly available for everyone to exploit, not merely to observe."

Since last year's announcement of the

chromosomal location of BRCA2, Stratton and his colleagues at the ICR have identified a region of the chromosome in which the gene is thought to lie. (Intriguingly, this region is also thought by some researchers to contain a gene — possibly also BRCA2 — involved in pancreatic cancer.)

Applying the technique of restriction fingerprinting to clones taken from an artificial chromosome library at Roswell Park Cancer Institute in Buffalo, New York, an ordered array of overlapping clones spanning this region was subsequently put together.

This has enabled sequence data for the region to be compiled by researchers at Washington University's Genome Sequencing Center and the Sanger Centre, part of the Wellcome Trust Genome Campus outside Cambridge, which is jointly funded by Wellcome and the Medical Research Council. On 23 November, the preliminary results were made available as an assembled shotgun sequence on the ftp sites of the two research organizations.

Three main research teams are perceived as being in a race to locate BRCA2. The one headed by Myriad Genetics, which many consider won the BRCA1 'race' largely due to the sheer size of its investment in automatic sequencing technology, is considered by many as the favourite. Close behind, however, are Stratton's team at ICR and its collaborators, as well as a consortium at Columbia University in New York which, like scientists at the Sanger Centre itself, has been investigating on the characteristics of chromosome 13.

Each of these three groups is thought to

have downloaded the eagerly awaited sequence data after it went on-line at 3pm last Thursday. But other groups, including one in Japan and another in Texas, are believed to be keen to take up the challenge.

Now the starting gun has been fired, it is up to the rival teams to see who can first identify the gene and the mutations that can cause increased susceptibility to breast cancer. "We have not had any preferential treatment," says Richard Wooster, a member of the ICR team that has localized the gene.

Sulston acknowledges that the decision to provide open access to the sequence — which still allows the team that finally identifies any mutations to the gene to apply for the appropriate patents — is an experiment whose outcome is uncertain. "We are taking a risk, and we do not know how the patent lawyers are going to respond," he says.

But he says the Sanger Centre wants to make a deliberate gesture in the controversial debate about how concern for proprietary rights should influence decisions about placing sequence data in the public domain. "We are campaigning for the notion that a sequence as such is 'pre-competitive' [information]," says Sulston.

"It will not be helpful to medicine if, by the year 2003, control of every single gene is tied up by one company or another for 20 years. That would be an enormous ball and chain. It seems obvious that, for the good of humanity, we should try to keep these things in the publicly exploitable domain."

The move has been welcomed by many scientists who expressed concern last year that the award of a broad patent to Myriad meant many others who had contributed indirectly to the BRCA1 discovery would receive neither recognition nor reward. "I am all for sharing recognition of this kind," says Bruce Ponder of the department of pathology at the University of Cambridge.

But others may have mixed feelings. The decision could backfire, for example, on British sponsors of the sequencing work if commercial returns from the eventual discovery of BRCA2 end up in the hands of a US company. Last year, for example, it was only after the NIH had filed an alternative patent claim on the BRCA1 discovery that Myriad and the University of Utah agreed that the names of two NIH researchers be included on a single, unified application. The NIH is now entitled to a share of the revenue from any commercial product that may stem from the BRCA1 patent.

The BRCA2 data released last week can be found on <ftp://ftp.sanger.ac.uk/pub/human/sequences/13q> and on <ftp://genome.wustl.edu/pub/gsc1/brca2>. **David Dickson**

IMAGE  
UNAVAILABLE  
FOR  
COPYRIGHT  
REASONS

James King-Homes/SPL

**Sulston: patent rush is 'counterproductive'.**

## UK court rejects appeal over hepatitis C kits

**London.** The medical supply company Murex Diagnostics has lost another round in its worldwide campaign to challenge the patent relating to diagnostic kits for the hepatitis C virus held by its rival, Chiron Corporation (see *Nature* 370, 493; 1994).

Earlier this month, the UK Court of Appeal dismissed an appeal by Murex against the decision of a patent judge, Mr Justice Aldous, in the summer of 1993, supporting claims by Chiron that Murex had infringed its hepatitis-C patent, and upholding the validity of the patent itself.

The decision has been welcomed by many patent agents as "good news for the biotechnology industry", as they claim that it helps lay to rest the impression created by several earlier decisions that UK courts are biased against biotechnology patents.

But the verdict is a blow to Murex, which had previously protested that Chiron's successful defence of a broad-based patent had prevented Murex from conducting research into alternative forms of diagnostic kits. The company is now considering whether to appeal to the House of Lords. **D. D.**



# Press panel recommends a unified budget for science

**Washington.** The US government should change the way it adds up its investment in science and technology (S&T), and prepare a specific S&T budget proposal each year for consideration by Congress, according to a report published this week by the National Research Council (NRC).

The report was prepared for the Senate Appropriations committee by a panel chaired by Frank Press, former president of the National Academy of Sciences and, prior to that, science advisor to President Jimmy Carter. It rejects the idea of a Department of Science, being backed by various Republicans in Congress, and suggests instead a stronger co-ordinating role for the White House's Office of Science and Technology Policy (OSTP).

The NRC proposes a new measure for science and technology spending designed to underpin a more tightly-integrated budget process. This 'metric' would exclude about \$30 billion which is currently spent on development and testing by the Department of Defense (DoD) and other agencies, leaving a total science and technology budget of roughly \$38 billion.

OSTP and the Office of Management and Budget (OMB) would, each year, lead a public process to set priorities within that budget, says the report. Congress would develop a way to examine the science and technology budget as a whole, it suggests, before it is split up and considered by individual appropriations subcommittees.

Senators asked for the report a year ago, when Tom Harkin (Democrat, Iowa), then chair of the Labor, Health and Human Services and Education appropriations subcommittee, wanted to know why there was no extra money for research at the National Institutes of Health (NIH), when the federal government was spending \$70 billion a year on R&D (see *Nature* 372, 5; 1994).

At the time, Press said that he hoped the report could match the influence of Vannevar Bush's influential *Science — The Endless Frontier* which, in 1945, first established the principle of substantial US government support for science in peacetime.

But where Bush offered specifics — including ambitious budget targets for his proposed National Research Foundation, which would emerge in 1950 as the National Science Foundation — the Press document

avoids numbers. Instead, it offers proposals for reforming the vast science-funding structure that has developed since Bush's day.

Marcia McNutt, professor of geophysics at the Massachusetts Institute of Technology (MIT) and a member of the NRC panel, says that the proposed reforms would mean "a larger role for OSTP and for the president's science advisor". Press, when he occupied the latter post, is remembered for having established a close working relationship with OMB.

The new 'metric' proposed by the panel would remove systems development work worth \$26 billion at the Department of Defense (DoD), \$3 billion at the National Aeronautical and Space Administration (NASA) and \$1.3 billion at the Department of Energy (DoE), from the government's conventional calculation of the money it spends on science and technology. This would leave the \$38 billion that constitutes investment in new knowledge.

The panel recommends that in the latter area, the President should present an annual budget, "including areas of increased and reduced emphasis". It concedes that recent administrations have already taken steps in this direction, but argues that they have not gone far enough. "Right now, the science budget is put together completely after the fact," says McNutt.

The panel asks Congress to "create a process" that would enable the President's science budget to be considered as a single entity by both the budget committees and the full appropriations committees. The Congressional Budget Office would also be asked to track the progress of the science budget through the appropriations process.

In rejecting the idea of a Department of Science, as proposed by Congressman Robert Walker (Republican, Pennsylvania), the panel argues for a diversity of funding sources. It says the Walker plan would fail to co-ordinate science across government, because it excludes both DoD and NIH.

At the same time, the NRC panel voiced scepticism about some technology programmes being pursued by the Clinton administration. The panel says it doubts that the Advanced Technology Programme is the best way to use scarce federal R&D dollars, and suggests that funds for cooperative research and development agreements (CRADAs) could be better spent elsewhere.

This week, Press has been briefing administration and Congressional staff on the report. The administration is likely to welcome its emphasis on an enhanced role for OSTP. But it is the response of Congress that will determine whether it has any lasting impact.

**Colin Macilwain**

## Environment agency is urged to focus on strategic priorities

**Washington.** The US Environmental Protection Agency's (EPA's) Office of Research and Development (ORD) last week released a draft strategic plan listing six high-priority topics for future research in the agency. The report recommends augmented programmes to study disinfectants in drinking water, airborne particulate matter, human health hazard assessment, ecosystem protection, endocrine disruptors and the use of new technology in pollution prevention.

According to Joseph Alexander, deputy chief of science at EPA, detailed research plans for each of these six areas should be ready by next summer. Some may receive additional funding in 1997, although fully developed research programmes are likely to wait until 1998.

The strategic plan calls for EPA to "focus research and development on the greatest risks to people and the environment, taking into account their potential severity, magnitude and uncertainty". The draft report, submitted to both the agency's science advisory board and the National Research Council, is part of an broad effort by the environmental agency to improve the quality of its science.

EPA has also asked the National Academy of Sciences to propose environmental research priorities for the next decade. Alexander says that the panel for that study, which is still being selected, could begin work by January.

In a foreword to the report, Robert Huggett, chief of science at EPA, writes that the choice of risk-based decision making principles for setting research priorities will "transcend economic and political changes". But that may be wishful thinking. Several Congressional committees have tried this year to prohibit EPA from conducting research in specific areas, including global warming and indoor air pollution.

The bills containing these provisions have yet to pass, and it is not certain how binding the restrictions will be if they do. But Congress will certainly have a say in whatever research priorities EPA proposes. In fact, EPA's funding bill, which is being voted in by Congress this week, requires ORD to report to the Senate on how and where it intends to spend its extramural research dollars.

The broadly worded strategic plan sets a number of ambitious goals for EPA science. These include improving the science of risk assessment, and integrating human health research with ecosystem science. But that work will require manpower, and Huggett says he is disturbed by language accompanying the EPA funding bill implying that the agency will have to cut its work force in the years to come.

**Tony Reichhardt**

IMAGE  
UNAVAILABLE  
FOR  
UNAVAILABLE  
COPYRIGHT  
REASONS

**Press: rejects plan for Science Department.**

Bachrach

# Oncomouse hearing ends up in confusion

**Munich.** An appeal hearing held to examine challenges to the first European-wide patent on an animal — Harvard University's so-called 'oncomouse' — broke up in disarray last week, when the opposition board of the European Patent Office (EPO), which should have given immediate judgement, found itself unable to answer some of the arguments raised against the patent.

The patent in question was granted in 1991 to Harvard on a mouse whose susceptibility to cancer is increased by the insertion of oncogenes into its genome. It had been challenged by a number of animal rights and environmentalist groups as conflicting with a clause in the European Patent Convention (EPC) that forbids patents on inventions considered to be contrary to public morality.

But after three days of listening to evidence related to 17 different appeals against the patent, the board, apparently thrown into uncertainty by arguments put forward by opponent groups at the oral hearing to back up their earlier written claims, unexpectedly drew the hearing to a close.

The board's declaration that the proceedings would continue only in writing angered the lawyers of both opponents and patent holders, who argued that the decision conflicts with the EPO's established procedures. No timetable was given for the written proceedings, or the likely date of a decision.

The EPO appears to have been embarrassed by the breakdown of its procedures during such a highly visible case. But its eventual verdict is still keenly awaited. Since the oncomouse patent application was filed with the EPO in 1985, the office has received more than 300 applications for patents on animals. Only three — including the oncomouse — have so far been granted. Decisions on the remainder are awaiting the outcome of the oncomouse case.

In initially granting the oncomouse

patent, the EPO, which was set up to carry out the procedures established in the EPC, interpreted the morality requirement to mean that benefits to humans must outweigh the harm to animals, and decided in favour of the value of the oncomouse as an animal model for human cancer.

But this interpretation has been challenged by various pressure groups, including in particular two British groups — Compassion in World Farming (CIWF) and the British Union for the Abolition of Antivivisection — and the mainly German group, Kein Patent auf Leben (No Patents on Life).

In addition to the claim that the oncomouse patent conflicts with the morality clause, these groups argue that, even within the EPO's own rules, the patent is invalid. One reason, they say, is its breadth, as the patent covers any mammal genetically engineered for increased susceptibility to cancer.

The groups argue, for example, that demands for the patent to cover an 'oncogiraffe' should be rejected on three related grounds. Firstly, the patent discloses insufficient information to allow an independent expert to create such a transgenic animal. Second, there is no moral justification for an 'oncogiraffe', as it could not be used as an experimental animal, and therefore could not be considered beneficial to humans. And third, there is no known industrial application for such an animal.

To the surprise of participants, who had expected a clear decision at the end of the week, the board of appeals appeared to indicate in the middle of last week's hearing that this claim had some legitimacy. It asked both sides for comments on a proposal that the patent should be restricted either to experimental animals or to rodents.

"This is the first indication we have had that the EPO thinks there is substance in the opponents' claims," says Peter Stevenson, a



**Not a squeak: protest at the patent hearing. No timetable has been set for the ruling.**

lawyer with CIWF. But, he adds, such a restriction would be only a partial victory for the opposition groups, whose main goal is that the patent should be revoked in full.

Another argument used by the opponents was that the oncomouse should be seen as an animal species, and thus as a collection of animal varieties. As 'varieties' of both animals and plants are excluded from patentability by the EPC, they claim, that in itself makes the oncomouse patent invalid. Lawyers for Harvard maintain that the two taxonomic levels can and should be distinguished — a position that has until very recently been shared by the EPO.

Now, however, this broader issue of the scope of the 'varieties' exclusion is under consideration by the EPO's highest legal authority, the Enlarged Board of Appeals, which is considering a patent granted to Plant Genetics Systems (PGS) of Ghent, Belgium, on a procedure for producing herbicide-resistant plants.

The PGS patent was partly revoked in February, when an EPO board of appeals ruled that it could not cover plants and seeds arising from the process (see *Nature* 374, 8; 1995). This was the first time that the EPC had been interpreted as meaning that plants themselves, as well as plant varieties, were not patentable.

But Paul Braendli, the president of the EPO, has asked for this decision to be considered by the office's Enlarged Board of Appeals, as it appears to conflict with previous rulings on the patentability of plants.

Braendli has also asked patent examiners not to use the PGS decision as a precedent before the higher board gives a final ruling. But in preliminary written comments on the opposition claims, published recently, the oncomouse board of appeal said it would not wait for the Enlarged Board of Appeal's parallel judgement on plants, which could take many months.

**Alison Abbott**

## Patently a matter of principle, not profit

**Munich.** As far as industry is concerned, last week's opposition hearing on the Harvard oncomouse patent (see above) was very much a battle of principle, rather than a defence of the profitability of the oncomouse itself, which does not appear to have been the commercial success originally expected.

Sales of the oncomouse were licensed to Dupont, the Delaware-based chemical company which subsidized the research leading to its development by scientists at Harvard University. Dupont took up the licensing rights to which it was entitled under its funding agreement with Harvard.

But a spokeswoman for the company

says that Dupont later changed its mind about the value of licensing and sub-licensing sales of the animal. "In order to bring therapeutics faster to the market there was more merit in having the animals made widely available," she says, adding that the company often gave away oncomice to university research groups free of charge.

The university will not say how much it has received in royalties from oncomouse sales. Meanwhile Dupont will soon be free of the whole issue. It is selling off its medical products division to concentrate on its core chemicals activities — and the sale will include its rights to the oncomouse.

**A. A.**

# CERN to open US talks on collider project

**Washington & London.** The director general of the European Laboratory for Particle Physics (CERN), Christopher Llewellyn-Smith, is to open talks with US government officials in Washington early next month about a possible US contribution to the proposed Large Hadron Collider (LHC).

According to officials of the US Department of Energy (DoE), the meeting will allow CERN formally to tell the United States what form it would like such a contribution to take. The US negotiating team, led by Martha Krebs, director of the DoE's Office of Energy Research, is not expected to bring its own proposals to the meeting, but to respond to those of CERN.

The US team will also include representatives of the National Science Foundation, the White House Office of Science and Technology Policy and the State Department. The meeting was to have taken place earlier this month, but was postponed because of the US government shutdown.

The meeting is the first of what may turn out to be a lengthy series. The DoE is cutting many of its own domestic programmes, and has little room for manoeuvre in supporting an overseas facility.

At the same time, however, the DoE budget for high-energy physics — the likely source of any LHC funding — has not been cut. A panel chaired by Sidney Drell, deputy director of the Stanford Linear Accelerator Center (SLAC) in California, last year reported that the United States could afford a substantial contribution to LHC starting in the 1998 financial year, when commitments to some domestic projects would

start to tail off (see *Nature* 369, 266; 1994).

Drell foresaw a US contribution of \$400 million between 1998 and 2003. He also said participation in the LHC would require the public endorsement of the president of the United States, which has not been forthcoming. At next month's meeting, Llewellyn-Smith is expected to lay out his proposals on how the United States could contribute to the LHC at a level commensurate with the expected participation of US physicists in experiments on the facility.

There seems to be general agreement that the bulk of any US contribution to the LHC would be spent in US laboratories, such as the Brookhaven National Laboratory and Fermilab. There is also confidence that the United States can be persuaded that its involvement in the project will provide some crucial 'added value' to the whole project.

Speaking at a meeting in Italy three weeks ago (see *Nature* 378, 226; 1995), Llewellyn-Smith said that the decision to build the LHC in two stages, with the final completion date determined by the levels of contribution from non-member states, offered an additional incentive for such countries to join. "We can now say to them 'If you come in, your contribution will allow us to get more value for money'," he said.

But there is also concern among CERN officials that, if any eventual US contribution is distributed between the costs of the main accelerator and the additional costs of experimental detectors in the same proportion as followed by most large European states — namely two to one — then the sug-

gested \$400 million could be insufficient.

In particular, the worry is that, if roughly two-thirds of the proposed \$400 million contribution goes towards the cost of building and operating the LHC, this could leave insufficient funds to meet the experimental needs of US physicists planning to use the collider's two detectors.

Krebs and Llewellyn-Smith will try to reach agreement on the acceptable form of a potential US contribution before next autumn, when the Clinton Administration has to prepare its budget for the 1998 financial year. But any agreement they reach will still have to be endorsed by Congress.

Given the strength of isolationist sentiment in the House of Representatives, the scars left by what is regarded as the failure of foreign governments to support the abandoned Superconducting Supercollider (SSC), and the relative novelty of being asked to participate in a major international project as a 'junior partner', that may well prove an even tougher obstacle than the negotiations themselves.

**Colin Macilwain & David Dickson**

## Stanford accelerator announces staff cuts

San Francisco. The Stanford Linear Accelerator Center (SLAC) in Menlo Park, California, is to shed between 60 and 70 staff jobs by next April, in anticipation of a \$7-million fall in the centre's operating budget for the fiscal year 1996. The cuts follow a decision by the Department of Energy to reduce SLAC's budget from \$107 million to \$100 million.

The lay-offs will reduce SLAC staff numbers from their current level of 1,350 to approximately 1,280. P. A. Moore, a spokeswoman for SLAC, says further redundancies depend "on what Washington has in mind". She claims that, although future reductions in staff numbers cannot be entirely ruled out, "we don't think they will be as severe as this". But some fear that the centre may eventually have to reduce its staff by up to 300 individuals.

SLAC officials have invited older staff to apply for early retirement to meet reductions intended primarily at reducing the operating overheads of the centre. But they also anticipate having to make some forced redundancies, as many of SLAC's older staff members took advantage of an early retirement scheme two years ago. "My guess is that we took most people at that cut," says Moore.

Nonetheless, Moore maintains that the lay-offs will not affect SLAC's present or future planned research programmes, including the construction of the B-factory that was approved last year. □

## French agency to protect visiting posts

**Paris.** The administrative council of France's Centre Nationale de la Recherche Scientifique (CNRS) last week unanimously endorsed the FF13.3-billion (US\$2.7-billion) budget decided by the government for next year, even though this will allow no real increase in support for laboratory costs and will cut the recruitment of young scientists.

The members of the council, instead, appeared relieved that CNRS has been shielded by the research ministry from budgetary reductions approved by parliament in October for all government departments. "The situation is quite difficult, but it could have been worse," says Edouard Henri Brezin, chairman of the council. "Seen from the point of view of the major financial crisis which we faced last year, we are beginning to see the end of the tunnel."

But many researchers remain concerned. Funds for laboratory support, which make up 20 per cent of the total CNRS budget, are being reduced by 7 per cent, to FF2.6 billion. Hardest hit by this reduction will be

newly created research teams and those working on novel research topics. Larger laboratories that are better placed to raise money from external sources will suffer less.

Similarly, the FF10.7 billion promised for salaries is insufficient to maintain recruitment of researchers at the same level as the rate of departure. There will be only 261 research posts available in 1996, compared with 324 this year.

Brezin says that the administrative council has been looking at ways of maintaining its ability to accept foreign researchers in CNRS laboratories without reducing its support for French scientists. Until now, foreign postdoctoral students have been recruited to research posts only for short periods. Taking account of low number of such posts now available, foreign researchers will in future be supported out of a laboratory's own funds. But since these are also being reduced, the effect may still be a reduction in this type of scientific cooperation.

**Catherine Tastemain**

# Temperature rises in dispute over costing climate change

**London.** Sir Crispin Tickell, one of the British government's leading advisers on environmental policy, has stepped into a fierce controversy about a United Nations (UN) report on the social and economic dimensions of climate change by suggesting that the use of cost-benefit analysis in a key chapter of the report is inappropriate.

Tickell, formerly Britain's ambassador to the UN, is now warden of Green College, Oxford. The author of a book on climate change, he made the comments in an exchange of letters with David Pearce, director of the Centre for Social and Economic Research on the Global Environment at University College, London, and one author of the forthcoming Second Assessment Report of the UN's Intergovernmental Panel on Climate Change (IPCC).

Tickell says cost-benefit analysis "should not be the basis — still less the sole basis — for making policy". But Pearce says Tickell's remarks are "wholly out of step" with government policy. "Being alone does not make you wrong," writes Pearce, a lead author of Chapter 6 of the section of the report prepared by IPCC's working group III. "But it ought to make you wonder if you have the basis for making such a judgement."

Tickell appears to disagree with the decision of the authors of Chapter 6 to use a technique for placing a value on loss of life whose implications are to assign a value in a developing country of one-fifteenth the value in the developed world, as it is based on a country's capability to pay for reduced risk. He describes the results of this method of calculation, known as 'value of statistical life', as "economists' artefacts of doubtful value and subjective character, with almost unlimited capacity to mislead".

But Pearce, questioning Tickell's understanding of the techniques of cost-benefit analysis, says developing countries cannot pay the same as the higher-income developed world. "The resources have to come from somewhere," he writes. "If, for example, they come from reduced foreign aid, we may kill more people than we save."

Pearce cites the lack of published literature exploring the impact of equal value statistical lives — an alternative method for placing a value on life — as another reason for not using Tickell's ideas as the basis for calculations in Chapter 6. "Our remit was to describe what the literature says, not to

rewrite it, nor do original research."

The exchange is likely to add fuel to the debate already raging about the methodology used by the authors of Chapter 6 to estimate damage from climate change. The debate has pitted the authors against government delegates to the IPCC, and divided the environmentalist movement over calls from one group, the Global Commons Institute in London, that the chapter should be withdrawn (see *Nature* 378, 119; 1995).

One author has suggested that this should be done on the grounds that a summary designed for policy-makers and written by a team of experts from different governments appeared to contradict the chapter. The summary does not contain quantitative damage estimates, on the grounds that such values are subjective. This author now appears to have withdrawn his objections. And the IPCC says the chapter will be included in the Second Assessment Report due to be approved in Rome next week.

However, a letter in this issue of *Nature* (see page 433), signed by 38 scientists including Sir Martin Rees, Britain's Astronomer Royal, and Hans Peter Duerr, director of the Max Planck Institute of Physics, says the chapter must go, on the grounds that a summary contradicting the chapter's contents violates IPCC procedures.

But not all environmentalist groups agree with this stance. The Climate Action Network, an alliance of green groups working in the field of climate policy, says this single issue should not be allowed to obstruct the greater goal of limiting the harmful effects of climate change. The chapter contains "a lot of information relevant to policy-makers", says Bill Hare, a climate policy adviser with Greenpeace International.

Hare also says the Second Assessment Report will not be the last word. The report will be discussed at a meeting of the Subsidiary Body on Scientific, Technical and Technological Cooperation before it is put to the Conference of Parties. Omitting Chapter 6 before it is even published, "would destroy the IPCC's integrity as an impartial body and open the way for vested interests to interfere", adds Hare.

P. R. Shukla, professor of management sciences at the Indian Institute of Management at Ahmedabad and a lead author for Chapters 8, 9 and 10 of the report, agrees. "We must not throw the baby out with the bath water," he says. But Shukla, whose chapters calculate the costs of slowing down greenhouse gas emissions, says the IPCC should arrange an interim review — "perhaps a supplement to the second assessment report" — that takes notice of more recent literature from the developing world. **Ehsan Masood**

## Call for more research into health effects of Chernobyl accident

**London.** The World Health Organization (WHO) last week called on scientists to step up their research into the effects of the Chernobyl nuclear accident that took place almost a decade ago in Ukraine in the former Soviet Union.

The call came during a four-day international conference on the effects on health of the accident. Speaking on the last day of the meeting in Geneva, sponsored by WHO and attended by 600 delegates from 59 countries, Wilfried Kreisel, WHO's executive director in charge of health and environment, said WHO "would want to see the volume and quality of medical assistance and scientific research increased".

Kreisel added that the scientific community would be doing a "disservice" if it failed to extract benefits for mankind out of "this monumental human tragedy".

But some scientists claim any apparent slowing down in research on the impact of the accident should be blamed not on 'Chernobyl fatigue' in the scientific community but on the unreliability of Russian data and on the large evacuation programme after the accident.

Last week's meeting heard evidence of an increasing number of incidents of psychological disorder among workers who dealt with the accident and its aftermath, as well as among those who moved away from areas of high contamination. Delegates were also told of an increase in thyroid cancers among children, but of "no statistically measurable increase in leukaemia".

Much of this information has been known for some time (see, for example, *Nature* 375, 365; 1995). But the identity of the radioactive agents responsible for increases in the incidence of thyroid cancer in children remains unclear. WHO acknowledges that identifying the agent should be a priority. Iodine 131 is one suspect, but when used in medicine it has not been shown to cause thyroid cancer.

WHO has begun to address some of the problems through its recently created Radiation Emergency Medical Preparedness and Assistance Network (REMPAN). The organization is also establishing a separate Chernobyl-related project in St Petersburg, Russia, to deal with problems encountered by accident recovery workers, and has called for more humanitarian aid for the region.

According to several delegates to the meeting, one useful outcome was agreement that the Soviet authorities, despite criticism from the West — caused largely by the obsessive secrecy of the Soviet regime — "did a pretty good job of treating the casualties".

E. M.

IMAGE  
UNAVAILABLE  
UNAVAILABLE  
COPYRIGHT  
REASONS  
COPYRIGHT  
REASONS

Universal Pictorial



# Decision time for Europe's hot rock project

**Paris.** Geothermal energy research in Europe faces a make-or-break point next month, when participants in the European Hot Dry Rock (HDR) project are scheduled to decide whether to build a prototype plant for electricity production at Soultz-sous-Forêts near Strasbourg, in France, or to abandon the research altogether.

Building the pilot plant would require around ECU37.5 million (US\$50 million) over five years from the project's backers, France, Germany and the European Commission. The decision whether to proceed will be taken after two separate reviews of the programme requested by the commission and the French ministry of industry.

The reviews are focusing on the economics of the high-technology 'closed-circuit' geothermal energy systems being developed at Soultz-sous-Forêts. Cost-effectiveness is the main barrier to developing the technology, with costs depending on factors such as borehole depth, the steepness of the temperature gradient with depth, and the ease of pumping water through the rock.

Long-term economic prospects depend also on obtaining a large enough heat transfer surface underground — for example, by exploiting natural fractures in the rock —



**Full steam ahead? Experimental facility may be developed into a prototype geothermal reactor.**

and on the number of years that a site can be worked before the rock cools down.

From a technical point of view, however, it seems generally accepted that the closed-circuit systems, in which water is injected into rocks that do not normally support water-flow and recovered as steam, offer a path to much wider exploitation of geothermal energy. At present this is restricted to areas with natural sources of steam, such as the Philippines, which produces around one-fifth of its electricity this way.

Roy Baria, a scientist with HDR, is optimistic that the reviews will be positive. In

particular, the project team, which has drilled 4 km into granite at Soultz, is confident of producing electricity in a pilot 10-MW plant at less than the critical threshold of ECU0.12 per kW hour.

Moreover, experiments carried out this summer suggest that the major technical difficulty of recovering all the injected water has been solved, according to Philippe Chartier, the scientific director of the French national environment and energy agency (ADEME), which is taking part in the programme.

But at least one European country is likely to be sceptical. Last year, the United Kingdom abandoned its domestic geothermal energy project in Cornwall, on which it had spent £40 million over two decades, following a review by the Department of Trade and Industry (DTI).

Britain also withdrew from the European HDR project last year after a review of its overall renewable energy research. According to one DTI official, the department believes that HDR is neither technically nor economically feasible in the UK in the foreseeable future. "We exhausted all the technical possibilities," he says, adding that the UK does not intend to reconsider its participation in the European HDR project.

But John Garnish, a European Commission official involved in HDR, argues that the DTI review was unfair, claiming that the department assessed the Cornwall project on the grounds of economics rather than as a research programme on 'rock mechanics'.

Garnish says he is reasonably optimistic that next month's decision will be positive, adding that the outcome will be critical for the future of geothermal energy in Europe. He points out that the technology is now mature, and that in the absence of a decision to pursue commercial development, there is unlikely to be support for further research.

The Soultz-sous-Forêts facility has a major advantage, in that the rock is much more fractured than the sites tested previously in the United Kingdom and the United States, making the project there technically easier and more efficient. But this feature is also a drawback, admits Chartier, as it limits the relevance of results obtained at the site to rock with similar properties, which, although common worldwide, are less abundant than rocks having few fractures. The Soultz programme is a "halfway step" towards the goal of wide exploitation of geothermal energy, says Chartier. "It's an apprenticeship."

If the participants approve the programme, building of the prototype would begin next year, and management handed over in 1998 to a consortium of utility companies, including EDF (France), Pfalzwerk (Germany) and ENEL (Italy), according to programme officials.

**Declan Butler**

## Renewable energy regains its funding

**Munich.** Europe's research commissioner, Edith Cresson, last week countered allegations that the European Commission had redirected money intended for renewable energy research projects to mainstream energy projects by agreeing to release sufficient funds to make up the balance between the two areas.

But she did not explicitly concede that the commission had been at fault in the way that it has administered funding of the five-year Joule programme into non-nuclear energy, which has a total budget of ECU967 million (US\$1.3 billion).

In the summer, the commission came under fire from the European Parliament after allocating only 40 per cent of the funds for Joule's first funding round, worth ECU191 million, to renewable energy projects, despite a parliamentary decision that such projects should receive 60 per cent of the total funds (see *Nature* **376**, 628; 1995). At the time, commission officials said they had not received enough strong applications in the area of renewable energy, and that the balance would be made up in future rounds.

An investigation by a parliamentary committee subsequently revealed that a senior commission official had over-ridden expert evaluations of some

projects. While accepting that the commission has the right to make such changes, the committee asked for a clarification of the grounds on which these decisions had been made, and complained that the commission had applied procedures that were not 'transparent'.

In replying to the allegations last week, Cresson said the commission remained committed to the rules agreed with parliament for the allocation of the Joule funds. She said a reserve list of projects submitted in the first round, giving priority to renewable energy projects and worth a minimum of ECU10 million, would be immediately drawn up. At least part of a programme's reserve list is usually funded.

Cresson also said the commission will put out a special call for proposals early next year. It will be aimed at renewable energy projects, with priority for solar energy, in order to make good the shortfall in funded projects.

Nuala Ahern, a member of the European Parliament who was closely involved in the investigation, says that parliament has accepted the response of the commission. "But we will continue to keep a close eye on how the commission operates," she says. **Alison Abbott**

# Deaths prompt calls for drug control review

**Tokyo.** A series of drug-related deaths that followed the prescription of drugs for non-approved applications has renewed calls for a review of the way Japan regulates drug safety. Critics say the government needs to establish an effective system for monitoring adverse drug reaction, and to put more resources into its drug approval process.

The calls for reform follow the release of a report from the Ministry of Health and Welfare (MHW) revealing that four premature babies died and three others were in a critical condition after they had been prescribed mefenamic acid for the treatment of arterial malformations.

In a separate incident, at least one patient died after being prescribed Danazole, an approved treatment for certain causes of infertility in women, for the treatment of idiopathic thrombosis (IDP). Danazole is recognized as an effective non-steroidal treatment for IDP. But critics claim that no effective watch was kept on its side-effects, as the MHW had not given official approval for this application under its 'Hokenteikyo' price-setting mechanism. Under this system,

the government sets a price for the drug and allows it to be paid for by the national health insurance scheme, which it administers.

A spokesman for the ministry's pharmaceutical affairs bureau denies that the safety aspects of approved and non-approved drugs are treated differently, and adds that it is common for drugs to be used without official sanction. Danazole, for example, has been used for treating thrombic conditions at a number of national and university hospitals since at least 1983.

But critics such as Masanori Fukushima, of Aichi Cancer Center in Nagoya, west of Tokyo, claim that official approval is being withheld from effective treatments because of under-staffing and other inadequacies in the drug approval process.

According to Fukushima, the MHW has only 20 full-time staff with a background in pharmacy dealing with drug applications, and no medical doctors; by contrast, the Center for Drug Evaluation and Research of the US Food and Drug Administration (FDA) has roughly 1,400 full-time staff, including 131 MDs.

As evidence of the Japanese system's failings, Fukushima points to the recent Sorivudine scandal, in which 15 people died within two weeks of the drug's approval (see *Nature* 369, 697; 1994) after they took it in combination with uracil-based anti-cancer drugs. Although the ministry blamed Nippon Shoji, the Osaka-based company that marketed the drug, Fukushima maintains that the MHW should have issued more explicit warnings of the danger.

There is also a long list of new drugs developed outside Japan that the MHW has so far failed to approve, including Taxol, a potent drug for treating ovarian cancer that was approved by the FDA in 1992, and Methotrexate, used in the treatment of small-cell lung cancer.

Fukushima claims that a fundamental problem in Japan's system of drug testing is the "lack of people in universities, in companies and in the MHW itself who possess the necessary training to oversee the clinical trial process". As long as this shortage remains, he claims, these deficiencies will persist.

**Stephen Barker**

## Safety dispute ignites over plans for UK nuclear sell-off

**London.** The controversy over plans to privatize Britain's nuclear industry intensified last week as opposing groups clashed before a parliamentary committee about whether the future private owners of nuclear power stations would be able to maintain acceptable levels of safety.

The Electricity Supply Trade Union Council (ESTUC), which represents many electricity workers concerned that they may lose their jobs under any privatization scheme, told a House of Commons select committee that privatization would compromise safety standards as companies sought increased profitability.

But a spokesman for the Health and Safety Executive (HSE) disagreed, telling the Trade and Industry committee that privatization may improve safety standards, as it will be in the industry's self-interest to generate and maintain public support.

The government has promised to privatize the most modern parts of Britain's £3-billion nuclear industry, including the country's seven modern advanced gas-cooled reactors, as well as its pressurized water reactor, Sizewell B in Suffolk.

Trade unions point to an accident in July 1993 at a nuclear power station at Wylfa, North Wales, as an example of the type of safety lapse they are concerned about.

The station's owner, Nuclear Electric, was fined £250,000 earlier this year. At the time, there were suggestions that a "commercially intense environment" leading up to privatization had contributed to the nine-

hour delay in shutting down the reactor after a crane fell into the reactor core.

But Chris Willby, deputy chief inspector of the HSE's Nuclear Installations Inspectorate, reassured the committee that lapses would not recur. "A lot has been learned since Wylfa," he said. "We must remember that it was two years ago."

Similar confidence was expressed by Jenny Bacon, director general of the HSE. She maintained that existing safety procedures would remain firmly in place, regardless of the ownership and structure of the nuclear industry. Bacon was confident that privatization, rather than being detrimental to safety, was likely to bring improvements.

ESTUC officials also fear erosion of the

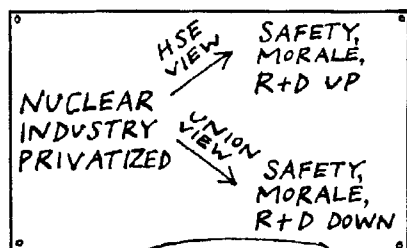
existing 'safety culture' within the industry because of low morale among staff, particularly those now working for Magnox Electric, the new publicly owned company that will retain the older Magnox reactors. These reactors will not be privatized as their total decommissioning costs of £8.5 billion exceed the revenue they could possibly generate during their lifetime.

Whereas Bacon told the committee that staff who were assigned to Magnox Electric had expressed optimism about their future careers in a company that still has 30 to 40 years to run, Tony Cooper, secretary of ESTUC, said that staff optimism stemmed from relief that the Magnox reactors were not being privatized. Such high morale, he added "can't last in a declining company".

There was also concern about the future of nuclear power research under privatization. The unions believe that short-term priorities may tempt the new owners of nuclear power to reduce long-term research spending, with serious consequences. "The research in the nuclear industry is concerned with safety and back-up. If the industry suffered as the coal-burning industry has, I think that would be a disaster," said Cooper.

But the HSE countered with the claim that research was stronger than ever. "Though we have powers to fund research if it is not met by industry, it is increasingly running research itself. We are seeing results two to three years before we would otherwise have done," said Willby.

**Ayala Ochert**



## Quis custodiet ipsos custodes?

SIR — Having attended the Conference on Plagiarism conducted by the National Institutes of Health (NIH)'s Office of Research Integrity (ORI), which was reviewed by John Maddox (*Nature* 376, 721; 1995), I can confirm that Charles W. McCutchen (*Nature* 377, 282; 1995) conveys quite well the flavour — and especially the scepticism — expressed by several in the audience about the integrity of ORI itself.

I too can attest to the presence of police — who were armed, to boot — and tampering with the transcript by ORI. Dr Ned Feder commented publicly on the presence of the police, adding that it was unheard of for armed guards to patrol the corridors of scientific meetings held at the NIH. Yet the transcript does not include Feder's remarks.

In an exchange between Dr M. Maureen Polsby, a former employee of NIH, and Dr Alan Price of ORI, Polsby rebuked ORI at length for its treatment of her complaint against NIH of sexual harassment and plagiarism. Her remarks and the response thereto were struck from the record.

Only one of my two comments was included in the transcript (without identifying me), the more damning of the two (where I took ORI to task for notifying the public only through announcements in the *Federal Register*) having been struck from the record.

Tampering with transcripts seems to be endemic with ORI. Affected participants in Federal Advisory Committee meetings of the Commission on Research Integrity, a sequel to the Conference on Plagiarism, have commented on excisions from transcripts of the meetings.

I suggest that there should be a congressional hearing on the operation of ORI.

**Burton L. Appleton**

927 East Taylor Run Parkway,  
Alexandria,  
Virginia 22302, USA

SIR — In a recent leading article about a conference on theft of ideas<sup>1</sup>, John Maddox denounced plagiarism as "the worst of bad behaviour"<sup>2</sup>. But is he too a purveyor of purloined goods? Among his other comments is this one: "The record of the meeting . . . clearly brings out the evident difference between scientists' and lawyers' conceptions of what constitutes plagiarism. To lawyers, the crime is akin to copyright violation, to scientists it is more like the commercial crime of 'passing off', which a manufacturer may seek to claim a distinguished brand name for inferior goods."

That analogy happens to be one I had drawn at a 1988 conference organized by

the Council of Biology Editors, during a session in which Maddox and I appeared on the same panel. My remark (as later formalized in the conference proceedings) was that: "adding the names of scientists to research for which they have no responsibility is false labelling: It misleads the consumer by promising a level of quality associated with a respected source"<sup>3</sup>.

Clearly Maddox did not use my words, but he does seem to have lifted my idea. In the process, however, he made two errors. First, I was referring not to plagiarism (putting one's own name on another's work) but to a type of 'gift authorship' (putting another's name on one's own work, to enhance its credibility). So Maddox had it backwards.

Second, Maddox attributed the "false labelling" view to scientists, apparently in the belief that they, not lawyers, understand the true nature of plagiarism. Although I have had some experience doing scientific research, I am now (and was in 1988) practising law. Perhaps I should be flattered that Maddox associates my ideas with the insights of scientists. I suspect that, in fact, he does not remember the origin of the analogy. (In cognitive science, this is called source amnesia; in journalism, as in research ethics, it's honest error.)

Still, we are left to brood (Maddox's word): is it plagiarism if someone steals your idea, but gets it all wrong?

**Barbara Mishkin**

Hogan and Hartson,  
Columbia Square,  
555 Thirteenth Street NW,  
Washington DC 20004-1109, USA

1. ORI/AAAS Conference on Plagiarism and Theft of Ideas, 21–22, 1993.
2. Maddox, J. *Nature* 376, 721 (1995).
3. Mishkin, B. in *Ethics and Policy in Scientific Publication* (eds Bailor et al.) 229 (Council of Biology Editors, 1990).

■ Mishkin raises a problem in modern etiquette. If you set out to persuade some of those who listen to you of a point of view, are those whom you persuade all plagiarists? And are the others the only honest people? I cannot recall her intervention in 1988 (no doubt a proof of "source amnesia"), but surely she should be glad, rather than irritated, that one member of her audience grasped (albeit incorrectly) the nub of what she said? — John Maddox, Editor, *Nature*. □

## Soros funds

SIR — The Soros scholarships to which you refer in a recent leading article (*Nature* 375, 264; 1995) exert a negative influence on the Russian educational system, despite George Soros's intentions. The selection of candidates for scholarships is made by elderly academics, so most scholarships have been awarded to

law-abiding lecturers in their declining years, who are unable to take account of modern scientific advances. What they teach their students was modern about 50 or so years ago. Receiving such a scholarship ensures the academic position of such people, and as a result the quality of teaching deteriorates.

**V. I. Rigin**

R&D Institute for Problems of Development  
of the Kansk-Achinsk Coal Basin,  
87 Kirensky-street,  
Krasnoyarsk 660041,  
Russian Federation

## Life assurance

SIR — In your report of a planned meeting between the Royal Society and the Institute of Actuaries (*Nature* 377, 375–376; 1995), it is stated that discussion will consider "how genetic analysis could be used to identify potential morbidity and mortality in groups of individuals". The life insurance premiums for such people can then be loaded to reflect the insurer's greater risk.

What if the analysis indicates exceptional longevity or resistance to certain diseases? Will the premiums then be concomitantly reduced? The insurers are strangely reticent on such a benign outcome.

**R. V. Harrowell**

6 Coach House,  
High Street,  
Codicote, Herts SG4 8XD, UK

## Price of stardom

SIR — I was amused by the eye-catching advertisement for Stratagene's Pfu DNA polymerase (for example *Nature* 12 October 1995, xii) featuring the Nobel prizewinner Kary Mullis. In the superposition photograph, Mullis changes from black to white lab coats — meant to indicate the mutating behaviour of Taq polymerase (sold by Perkin-Elmer) but liable to be misinterpreted as a 'turncoat' metaphor for promotion of Pfu at the expense of his former employer's product.

I wonder if this advertisement signals a new age in science commercialism in which scientist superstars will augment their modest incomes through advertising hooks much like vastly better paid (and more famous) star athletes? Unlike Charles Darwin and Gerhardus Mercator, whose names have been appropriated by two young biotechnology companies, Mullis (assuming he was paid) is cashing in on stardom in his own lifetime.

**Alexander Kamb**

Myriad Genetics Inc.,  
390 Wakara Way,  
Salt Lake City,  
Utah 84108, USA



# Economics of climate change

SIR — Further to your news report on the economics of climate change (*Nature* 378, 119; 1995), I write on behalf of myself and the undersigned\*. We note that the Intergovernmental Panel on Climate Change (IPCC) is now due to approve for publication its full Second Assessment Report (SAR) at its plenary meeting in Rome on 11–15 December.

The assessment by Working Group Three (WG3) of the "Social Costs" of climate change (or "damages") will be included in the SAR. This contains the now notorious 15:1 mortality costing between rich and poor people in developed and developing countries. This largely explains why the overall damage figures cited in the chapter (1.5–2 per cent of gross world product) are so low.

Both the global and the regional damage figures are widely regarded as unsafe, so much so that the Summary for Policy-Makers (SPM) of the "Social Costs" written by the governmental representatives at the last WG3 meeting omits reference to these quantitative damage results altogether.

In fact, rather than being a 'summary' of the chapter, the SPM largely concentrates its comments on how much higher the damage results would have been had non-discriminatory methods of valuation been used. This has produced a marked inconsistency between the chapter and its summary, which the authors of the chapter themselves have confirmed.

If IPCC puts its imprimatur on this material by publishing it, this unsafe and discriminatory data will become official advice to the UN negotiating process for at least the next five years. This would give

a disastrously wrong signal at a time when it is becoming increasingly clear that serious policy measures to arrest climate change are now required and when the political tensions over the "differentiated responsibilities" in this task are increasing as well.

Moreover, if the IPCC goes ahead and publishes in these circumstances, it will violate its own procedures. These clearly state that approval of the SPM signifies that it is "consistent with the factual material contained in the full scientific and technical assessment," and this is clearly not the case.

In these circumstances, IPCC's reputation for procedural correctness and consensus-building around scientific accuracy will be permanently compromised. Consequently we urge the rejection of the "Social Costs" chapter in the report.

**Aubrey Meyer**

Global Commons Institute (GCI),  
42 Windsor Road,  
London NW2 5DS, UK  
Fax +44 (0)181 830 2366  
e-mail [saveforests@gn.apc.org](mailto:saveforests@gn.apc.org)

## Finding fault

SIR — The intensity and duration of ground shaking at a site depends not only upon the magnitude of the earthquake but also upon its distance; therefore the statement that the Very Large Telescope being constructed on Mount Paranal in northern Chile was designed to withstand earthquakes of at least magnitude 8.5 is incomplete.

The building that will house the telescope is under construction and the earthquake of 8-magnitude on 30 July "seems to have caused little damage ... although damage to the shock-absorbing columns will require several weeks to repair" (*Nature* 376, 542; 1995). A magnitude 6.5 earthquake close to the site can produce ground shaking much more intense, but of shorter duration, than a magnitude 8.5 at a greater distance. The epicentre of this earthquake was 130 km south of the site of the telescope so it raises questions about the seismic design of the building, and of the telescope itself. If it has not already done so, the European Southern Observatory would be well-advised to seek the advice of an expert in earthquake engineering and strong ground shaking, as well as a geologist who can identify potentially active faults in the vicinity.

**George W. Housner**

211 Thomas Laboratory,  
California Institute of  
Technology,  
Pasadena, California 91125, USA

## Theism and science

SIR — I agree with Walter Gratzer's comment in his review of John Carey's *The Faber Book of Science* (*Nature* 378, 111–112; 1995) that theology can hardly "be regarded as a science", although it is none the worse for that, as neither can poetry or music. But Gratzer's further claim that "[t]he cast of mind that draws scientists to their profession is the antithesis of that which predisposes towards religion" is inaccurate. Apart from the well documented contributions of theists to the scientific enterprise, in particular since the founding of the Royal Society in the seventeenth century, there is clear evidence of a contemporary 'selection pressure' so that Christians tend to be more abundant in the sciences than the arts.

This has often drawn comment here in Cambridge, where scientists are far more abundant in the packed central city churches than those from the arts. A possible source of this 'selection pressure' is the shared scepticism of both Christians and scientists towards the more extreme forms of relativism promoted by post-modernism, and their common 'realist' stance which maintains that neither scientific nor religious knowledge is a merely social construct.

It would be of interest to know whether this disparity in the representation of Christians in the sciences rather than the arts is a Cambridge phenomenon only, or whether it reflects some wider trend.

**Denis R. Alexander**

Babraham Institute,  
Babraham Hall,  
Babraham,  
Cambridge CB2 4AT, UK

## Gloves off

SIR — In a recent leading article "Time to ban British boxing" (*Nature* 377, 561–562; 1995), you correctly predicted the reaction of the House of Commons to suggestions of a ban. In particular, the arguments for a ban on the grounds of the neurological damage to boxers and the degrading effect on society, were insufficient to counter the libertarian views of the supporters of boxing.

As boxing is not to be banned, perhaps the neurological damage to boxers could be limited by banning the use of gloves or any other protective device and declaring the winner to be the first to draw blood. This would result only in relatively superficial cuts and bruises instead of the present prolonged battering of the skull, with its consequent trauma to the brain. Promoters would also benefit from the increased audiences.

**Andrew J. Wilson**

49 Avenue du Mistral,  
1200 Brussels, Belgium

\*Tom Wakeford (Biology, York); Nigel Woodcock (Earth Sciences, Cambridge); Gerald Leach (Stockholm Environment Institute and IPCC lead author); John Whitelegg (Ecologica UK); Ben Matthews (East Anglia); Maneka Gandhi (ex-Environment Minister, India); Vandana Shiva (Centre for Science and Ecology); Bittu Sehgal (Sanctuary Magazine); Medhar Pakkar; Claude Alvarez (Third World Network, India); Peter Newell (Environmental Politics, Keele); David Smernoff (Bay Area Action, California); Ernst von Weizsäcker (Wuppertal Institute); Christine von Weizsäcker; Hans Peter Dürr (Max Planck Institute for Physics); Freda Meissner-Blau (ECOROPA); Charles Levenstein (Massachusetts and New Solutions); Brian Moss (Environmental and Evolutionary Biology, Liverpool); George Marshall (Earth Action Resource Centre); John Barkham (Environmental Sciences, East Anglia); Nicholas Hildyard (The Ecologist); Rudolf Lippe (World Decade for Cultural Understanding, UNESCO); Hartwig Spitzer (Center for Science and International Security, Hamburg); Hans Jürgen Fischbeck (International Network of Engineers and Scientists for Global Responsibility); Heinrich Volkert (Protestant Church of Germany); Simon Naylor (Geography, Keele); Martin Rees (Kings College Cambridge); Philip Webber (Scientists for Global Responsibility); Michael Benfield (Town and Country Planning, University of Newcastle upon Tyne); Wolfgang Sachs (Wuppertal Institute); Angelika Zahrt (BUND/ Friends of the Earth, Germany); Greg Samways (Environmental Sciences, East Anglia); Susan George (Transnational Institute, Amsterdam); Andrew McLellan (Institut für Umweltwissenschaften, Zurich); Philip Smith (International Network of Engineers and Scientists for Global Responsibility); Christian Azar (Chalmers University of Technology, Göteborg); Kari-Erik Eriksson (Theoretical Physics, Göteborg); P. R. Shukla (Indian Institute of Management, Ahmedabad and IPCC WG3 lead author); Tim Lenton (Environmental Sciences, East Anglia, and Plymouth Marine Laboratory).



# Aversion to xenotransplantation

**SIR**—A recent news item in *Nature*<sup>1</sup> reported that experts regard xenotransplantation as a new medical field with explosive growth potential. They foresee a

Opposition to xenotransplantation is not limited to acute care nurses, but is also expressed by others, including animal rights activists, actual transplant recipi-

commence clinical trials will inevitably be based on the experimental data.

**David White**

*University of Cambridge,  
Department of Surgery,  
18 Trumpington Road,  
Cambridge CB2 2AH, UK*

## ATTITUDES TO XENOTRANSPLANTATION

|                                                                                             | Strongly disagree |               |              |                |               | Strongly agree |               |
|---------------------------------------------------------------------------------------------|-------------------|---------------|--------------|----------------|---------------|----------------|---------------|
|                                                                                             | 1                 | 2             | 3            | 4              | 5             | 6              | 7             |
| "I would accept an organ from a species closely related to man (e.g. baboon or chimpanzee)" | 849<br>(51.1%)    | 156<br>(9.4%) | 90<br>(5.4%) | 254<br>(15.3%) | 121<br>(7.3%) | 89<br>(5.4%)   | 104<br>(6.3%) |
| "I would accept an organ from a species distant to man (e.g. pig or sheep)"                 | 880<br>(53.3%)    | 126<br>(7.6%) | 85<br>(5.1%) | 248<br>(15.0%) | 110<br>(6.7%) | 83<br>(5.0%)   | 119<br>(7.2%) |

The table shows Likert scale frequency and percentage responses to two questions about attitudes to xenotransplantation.

future in which animals would routinely provide bone marrow, hearts, lungs and kidneys for humans, and estimate that 50,000 pig hearts and 40,000 pig kidneys could be transplanted into humans each year.

While these views may reflect the general attitudes of those involved in transplantation and xenotransplantation research, the views of other health professionals suggest differently.

Our analysis of a study of attitudes towards organ donation of 1,728 acute care nurses in 59 Australian public hospitals shows that the majority of these nurses are opposed to xenotransplants. Using a seven-point Likert scale, nurses were asked to grade their feelings towards the two statements: "I would accept an organ from a species closely related to man (for example, baboon or chimpanzee)", and "I would accept an organ from a species distant to man (for example, pig or sheep)".

Approximately two-thirds (65.5 per cent) of nurses disagreed with the use of primates as organ donors, 19.0 per cent agreed, 15.3 per cent were undecided. Their responses to the use of pig or sheep organs were closely similar: 66.0 per cent were opposed, 18.9 per cent agreed, 15.0 per cent were undecided. The strength of disagreement with the use of organs from these species is apparent from the table.

ents<sup>2</sup>, potential transplant recipients (I. F. C. McKenzie, personal communication) and the general community<sup>3</sup>.

The current huge expenditure of money and time on xenotransplantation research will be of little value if people are unwilling to accept animal organs. Reemstma<sup>4</sup> suggests that when biological barriers to xenotransplantation are overcome, the ethical problems will recede in proportion to the degree of success achieved. However, even if immunological barriers can be overcome, and fears about risks of transmitting donor viral infections allayed, the social unacceptability of clinical xenotransplantation may still prevent it from resolving the present worldwide shortage of organ donors.

**Paula J. Mohacsí**

**Charles E. Blumer**

**Susan Quine**

**John F. Thompson**

*Department of Public Health  
and Community Medicine,  
University of Sydney, NSW 2006,  
Australia*

**SIR**—Your News story about our work with transgenic pigs<sup>5</sup> raises the important question of timing of initial human studies: what primate data would justify a clinical trial? Some people say that 30-day survival would be sufficient to justify using xenografts as a 'bridge' to transplantation. Others would insist on extensive prolonged survival studies. Clearly, reasonable survival targets must be established somewhere between these two extremes.

Because of the rapid advances in long-term concordant xenograft survival, there is a realistic possibility that our studies on long-term immunosuppressive strategies could, over the next 12 months, deliver such targets. It is therefore appropriate that plans for a clinical programme dealing with such issues as disease control, ethical considerations and regulatory affairs should be initiated as soon as possible. The final decision on when to

**SIR**—With the increasing demand for transplant donors, research continues on the possibility that animal organs could be used. This research often centres on pigs (see, for example, ref. 6), because it is believed that their size and general physiology matches that of humans. What never seems to be discussed is the fact that pigs have quite short life spans. Indeed, amongst mammalian species of comparable size, pigs seem to have the shortest life span<sup>7,8</sup>. This is about 20 years for wild pigs and significantly shorter for domestic animals, which have been selected for particular physical traits. Thus much research on ways and means of overcoming the recipients' rejection mechanisms might be completely foiled by the fairly rapid ageing of the transplanted heart. This would be particularly serious for the most deserving recipients, namely those who are not too old.

This presupposes, of course, that the ageing of the pig heart is intrinsic to its cells and tissues, and is not determined by some extrinsic organism-based mechanism. There is much evidence that ageing at the cellular and tissue level is indeed intrinsic<sup>9</sup>. This includes transplant experiments. For example, it was shown many years ago that mouse skin transplanted to isogeneic recipients became senescent irrespective of the age of the recipient<sup>10</sup>. Similar results were obtained with the transplantation of rat mammary tissue<sup>11,12</sup>.

**Robin Holliday**

*CSIRO Division  
of Biomolecular Engineering,  
PO Box 184,  
North Ryde NSW 2113,  
Sydney, Australia*

## Bee versus hornet

**SIR**—Masato Ono *et al.* say (*Nature* 377, 334–336; 1995) that the European honeybee introduced into Japan cannot defend itself against the Japanese giant hornet. Beekeepers in Japan should therefore perhaps consider installing at the entrance to their hives a device similar to the 'queen excluder' used by British beekeepers. The marked difference in size between hornet and bee might make this a practical possibility, not only for the protection of *Apis mellifera* but also for that of *A. cerana japonica*.

**H. Heyworth**

*20 Bournside Road,  
Cheltenham,  
Gloucestershire GL51 5AH, UK*

1. Lehrman, S. *Nature* **376**, 8 (1995).
2. Okada-Takagi, M. & Williams, T. *Bioethics News* **12**, 12–30 (1993).
3. *The American Public's Attitudes Toward Organ Donation and Transplantation Conducted for the Partnership for Organ Donation*, Boston, USA (Gallup, 1993).
4. Reemstma, K. *Transplant Proc.* **22**, 1042–1043 (1990).
5. Dickson, D. *Nature* **377**, 185–186 (1995).
6. Hayssen, V., van Tienhoven, A. & van Tienhoven, A. *Asdell's Patterns of Mammalian Reproduction* (Cornstock, Cornell Univ. Press, Ithaca, 1993).
7. Nowak, R. M. *Walker's Mammals of the World*, 5th edn (Johns Hopkins Press, Baltimore, 1991).
8. Holliday, R. *Understanding Ageing* (Cambridge Univ. Press, 1995).
9. Krohn, P. L. *Proc. R. Soc. B* **157**, 128–147 (1962).
10. Daniel, C. W. *et al. Proc. natn. Acad. Sci. U.S.A.* **61**, 53–60 (1968).
11. Daniel, C. W. & Young, L. J. *J. Expl. Cell Res.* **65**, 27–32 (1971).
12. Daniel, C. W. & Young, L. J. *J. Expl. Cell Res.* **65**, 27–32 (1971).

# The prevalent distrust of science

John Maddox\*

**An account of the now common distrust of science, an explanation of its origins and a suggestion of what may, and indeed, should be done about it.**

DISTRUST of science is still alarmingly prevalent, which conflicts with reasonable expectation. Is not the century now drawing to a close most of all remarkable for the technology that now fills our world and for the understanding of that world that has been won since, say, the discovery of the electron in 1897?

During that long period, the improvements in the human condition have been immense. People will make their own list of the innovations that stand out in their minds — aircraft, radio communications, computers, drugs and antibiotics, now computers. In general, science and technology have helped to make us healthy, wealthy and wise in a manner and to a degree not foreseen except by a few visionaries such as H.G. Wells.

The assertion about health is evident in the now standard expectation of life in the rich countries of the world: three score years and ten has become the benchmark of performance. Another sign of the improvement is the large proportion of the national wealth spent on health care in the rich countries — 10 per cent of gross domestic product (GDP) is the standard figure. Another sadder indicator is the continuing disparity between the rich countries of the world and the poor ones, which spend much less and where people live less long.

The wealth that we enjoy is equally apparent. The standard measure is the number of US dollars of GDP per head of population. But we often forget that these numbers have a qualitative dimension. We eat well, almost all of us are comfortably housed, we travel the world, use the telephone frequently and look forward to increasing use of a whole new range of communications services. The benefits that flow from these developments are not just material; they are the source of the civility to which advanced societies aspire. They are also the source of the common conviction that we belong intimately to an increasingly interconnected world.

We are also wiser. Nobody disputes that we are smarter than were our forebears at the turn of the century. For example, despite the poor reputation of the economists, we now understand how industrial economies function. We are also smart enough to anticipate problems

that lie ahead, global warming for example. But are we wiser in a deeper sense? The understanding of the world won in the past century has given us a much better opportunity than before to understand our place in nature, and to understand ourselves.

Take, for example, the question of where the ingredients of our world came from. It's now just 40 years since Hoyle, Fowler and the two Burbidges showed that the elements of which the world (and we) are made are synthesized in stars essentially like the Sun. We understand why some elements, aluminium for example, are abundant and others such as uranium are not. That's no longer magic, but nuclear physics.

Similarly with the phenomenon of life. Watson and Crick's structure of DNA, published in 1953, is not just self-evidently a recipe for genetic inheritance (because of its double-stranded nature) but is also the means by which the chemical processes in every cell in every living thing function as they do. So the old question of whether there is something distinctive about living matter has become unfashionable. Life is chemistry. And vitalism is dead.

That's wisdom in a true sense, and it's invaluable. The better we understand the world and ourselves, the better we are able to look out for our survival. What is now being learned of human genetics (a product of what is known of DNA) will be dramatically reflected in the health of our populations in the decades ahead. But we have also been helped to understand that there is nothing magical about the ingredients of our world or about life.

It must have been the same when neolithic people first built those stone circles — Stonehenge is the best example in Britain, perhaps anywhere — whose purpose was evidently to construct a kind of calendar of the seasons. For one thing, they knew when to plant their crops, so that their survival was better assured. But they would also have learned something about the regularity of the Sun's movement across the sky, year on year, that would have dispelled at least some of the superstition surrounding that important question.

So why, given all these benefits of health, wealth and wisdom, to which science has made such important contributions, does there persist the deep distrust of science we see around us?

## Claims

The standard answer is that science and scientists have in the past made exaggerated claims of what innovation will do for the world at large, so that scientists are no longer trusted. There is something in that charge, but what are called the 'false promises' do not account for the continuing distrust of science.

Nuclear power is an illustration. There was great excitement about the peaceful uses of nuclear energy in the late 1950s and early 1960s, shared by both the technical community and others. We were told that energy shortages would be abolished for good. Astonishingly, in the autumn of 1955, the British government adopted a plan to build no fewer than four different types of nuclear reactors in the succeeding decade, culminating in the fast reactor that would generate, or 'breed', more nuclear fuel than it consumed. Then, when the dream of nuclear fusion came along, people worried aloud whether the cost of electricity would be determined by nothing more substantial than the cost of installing electricity meters to tell how much power individuals had consumed.

In that heady period, the promise of nuclear power was not exaggerated, but the technical difficulties of making use of it were made light of. The difficulties — safety, waste disposal, the exposure of people to radiation during the reprocessing operation — were all clearly identified. But when so many apparently insuperable problems in the building of reactors had been successfully solved, hardly anybody stopped to think that the technical difficulties of the fuel cycle, or the cost of solving them, would be a brake on the exploitation of nuclear energy. Then, nearly ten years ago, there was the accident at Chernobyl.

I belong to that now-small group of people who believe that nuclear power stations can be built and operated safely, that the disposal of radioactive waste is a soluble (indeed, a solved) problem, and that it would be prudent to plan for the building of nuclear power plants in large numbers if, in the years ahead, we discover that global warming is a serious threat. But that is some decades ahead.

What went wrong with nuclear power was not the promise but the business plan. Especially in Britain since the Second World War, we've had ample proof that turning a bright idea into a practical

\*Based on the text of the Voltaire lecture to the British Humanist Association 18 November 1995.

application is usually more difficult than is believed at first. Take the Comet aircraft, caught out by metal-fatigue, and the military aircraft called the TSR2, which was designed to follow ground-contours when travelling at very high speed, but which never succeeded in doing so. But not every project has been a disappointment. The jet engine in the end made new kinds of air travel possible, from the microchip has sprung the computers and the telecommunications services now changing our lives, and there has been a torrent of useful pharmaceuticals in the past few decades. For most people, the pluses far outweigh the minuses.

Technical people nevertheless have three lessons to learn from the recent history of innovation. First, they must be ready to moderate their enthusiasm for new ideas with an awareness of the difficulties of turning them into reality. Second, the difficulties may not be technical at all, but social; the brightest ideas (thermonuclear fusion, perhaps) may prove to be uneconomic, others may be practicable but unacceptable for other reasons. Finally, the enthusiasm of researchers for their discoveries can easily appear as a kind of triumphalism. The nuclear power saga of the 1950s may be one illustration, molecular genetics is at risk of becoming another.

But none of that is a sufficient explanation of the distrust of science among people whose backgrounds and education would suggest that they should know better. In the grown-up world, after all, people know that well-intentioned projects repeatedly come to nothing in all kinds of circumstances. Look at the attention lavished on the plight of the developing world, and the present condition of much of Africa. There's also the British government's failure to solve its problems in Northern Ireland. It's not just in science that high promise is denied by cruel circumstance.

## Roots

The general distrust of science has other and more primitive roots. To the extent that science and its applications bring improvements in our lot, they also imply change, and change is never welcomed for its own sake. Then this knowledge whose benefits I've been extolling is often unwelcome; it shifts responsibility from nature onto people's own shoulders and is often an uncomfortable challenge to deeply held beliefs. A further complication is that science as often presents people not with the dogmatic opinions for which it is often blamed but with uncertainties.

Uncertainty has become a great bugbear; it turns up all the time. There's the global warming issue, for example. It is a fair statement that there's an  $x$  per cent chance that the temperature on the surface of the Earth will increase by 1 degree

centigrade in about 50 years. In my opinion,  $x$  is about 80 per cent. Quite properly, there's still a vigorous debate about the accuracy of the prediction; there are even some researchers who believe that mechanisms that offset the simple greenhouse effect will eventually prove the prediction wrong. But that does not diminish the force of the conclusion that there's a high risk of a serious crisis in the next half-century.

That is why governments were persuaded to sign the Rio Treaty in 1990. There are many valid criticisms of the manner in which that was done, and of the adequacy of the steps since taken, but there can be no dissent from the view that these arrangements are a step in the right direction. But that is not the impression one gets by talking to civil servants or business-people, many of whom are engaged in planning projects likely to come to fruition only over several decades. They prefer to deny what probably lies in store, at least "for the time being", saying that they would prefer to wait for a scientific consensus. By then it may be too late.

## Anti-science

The character of the intelligent person's distrust of science is typified for me by something that happened soon after I first became editor of *Nature* in 1966. At dinner one night, I met a woman who was one of Britain's most distinguished architects (sadly, she's now dead) who, hearing of my new job, said laconically, "I suppose that you must be one of those frightful darwinists, then!" What she meant, it turned out, was that the theory of evolution had no place in her work and that the programme to understand how people and other species had evolved from earlier life-forms could only undermine people's sense of their own dignity and self-respect, and sap their will to aspire to better things. I never found out whether my acquaintance was a religious person.

Nearly 30 years later, I would have made a more robust defence of the "frightful darwinists". I would have argued that the comparison of the genome of human beings with those of the Great Apes and other primates may soon make it possible for us to understand the evolutionary history of the human race. Which came first, bipedalism or language? Exactly what were the adaptations that brought about the transition from *Homo erectus* to *Homo sapiens*? And when that history has been reconstructed, we shall, I suspect, be even more conscious than we are now of the natural accidents that shaped our evolutionary history, of our continuing fragility as a species and of the need that we should act responsibly in what Adlai Stevenson once called "the care and maintenance of a small planet".

That is anti-science by denial, or culti-

vated ignorance. As anti-science goes, it may even seem a benign manifestation of the phenomenon. The trouble is that there is a lot of it about. And while people may properly choose not to encumber their heads with information and insight that they consider to be irrelevant to their lives, the process of deciding what is relevant is evidently an opportunity for bias. People can and do elect to hide from the unpalatable, the uncomfortable and the subversive. In the process, they cheerfully misrepresent what science or scientists are saying. The distrust of science is really intentional obfuscation of what science is about.

The threat of global warming is an unpalatable piece of science from which people readily seek escape. So too is the need for sympathetic and constructive action to contain the growth of the world's population; the carrying capacity of the Earth is probably much greater than the present population, but the economic and social costs of letting it grow without restraint are much greater than we or our descendants can afford to pay. Neither of these forms of denial is all that benign.

Those are examples of passive denial, when people prefer to sit on their hands than to respond. There is also active denial. That is where the creationists, latterly calling themselves 'creation scientists', come from. They have two lines of argument: there are gaps in the fossil record (which everybody knows and most people expect), meaning that evolution is not proven; and they have come across phenomena, ancient zircons perhaps, in which the distribution of isotopes is more consistent with a recent than a distant age. (There is also the claim that the Universe was created 6,000 years ago in such a way that we have the illusion that it is really very old, but even the creationists seem unconvinced by that.)

In the second line of argument, their denial consists in denying the validity of all the evidence supporting an age of the Universe of about 10 billion years. The grounds are that their few alleged exceptions must be accorded at least equal validity. The opponents who take them seriously engage in endless arguments on the theme that the observations the creationists plead in their aid may be exceptional because the phenonema or objects observed may have had an exceptional history.

It is true, of course, that there are some fields of science in which a single discordant datum can bring a whole theory crashing down. If somebody could find a single exception to the assertion of Fermat's last theorem, for example, the business of trying to prove the theorem, which has occupied mathematicians for two centuries, would grind to a halt. But the observational sciences are not like that.

Discordant evidence is usually found to have an explanation.

In that sense, scientific creationism is not only a denial of the evidence but a misrepresentation of the scientific process. There are, of course, frequently occasions when people disagree about the explanation of the same set of data. Then, the argument turns on the endeavours of people honestly trying to make sense of it. In Britain, creationism is no longer a big issue. Even in the United States, it seems to be making little headway except in the southern states. But it remains a disturbing precedent.

Here is an illustration. A few weeks ago, at a meeting of Unesco's Commission on Ethics in Biomedicine, there was a panel of parliamentarians gathered to discuss the legislative position on genetics and genetic manipulation in their countries. A woman member of the German Bundestag, and a representative of the Green Party, spoke clearly and intelligently and said this: "You must understand that we Greens believe that to represent the nature of human beings by a description of their genes undermines their dignity as human beings. We shall oppose in the Bundestag any legislation that condones research in human genetics."

This implacable position is arresting. It also succeeds in misrepresenting the position of the research community. Broadly speaking, geneticists themselves are deeply suspicious of genetic determinism — the assertion that a person is determined almost exclusively by the genes there happen to be in his or her genome. To their credit, geneticists have also been among the first to draw attention to the respects in which the rapid development of their field is likely to create social problems, chiefly by the use of genetic diagnosis as a basis for discrimination between individuals, mainly in employment and insurance. But evidently the geneticists will win no credit from the German Greens for their perceptiveness. The danger, for the wider community, is that sensible legislation will be held up and those who might benefit from it will have to wait a little longer.

There are more subtle ways of misrepresenting science. We have all heard people in general proclaim that, while they "do not understand" quantum mechanics, they do know that quantum mechanics shows that mechanical processes are not nearly as well defined as Newton would have had us believe. Heisenberg has a lot to answer for. The same goes for the recently fashionable concept of chaos, often taken as a licence for the belief that nothing is certain any longer. But both sets of phenomena are deterministic in the philosophical sense. The state of a quantum or a chaotic system at a particular time depends only on its state at an earlier instant and the laws of physics.

That the outcome of such an evolution may not be a single state but, rather, may be one of several does not mean that the Universe is unpredictable, but simply that, within restrictions such as that the total amount of energy is conserved, some of its elements behave like that.

The reason why this line of speculation is now so fashionable is not far to seek. Since the time of Descartes, the world at large has been uneasy with the notion of the Universe as a machine whose fate may eventually be determined by calculation. People do not like the thought that there may be elements of our fate over which we have literally no control. Yet there is nothing in quantum mechanics or chaos theory that puts God back into mechanics. The laws of mechanics, which have been amended once since Newton's time, may have to be amended yet again if people manage to make a bridge between gravitation and quantum mechanics, but there is nothing to suggest they are other than deterministic, one state of affairs following from its precursors.

### Understanding

So what is to be done about the prevalence of distrust in science? Although the 'false promises' charge may not be the wellspring of distrust, giving hostages to fortune by over-optimism must be avoided at all costs. Perhaps science should be more aware that technical developments rarely become fully fledged technology without the intervention of other fields of learning. The social sciences just now may be under a cloud, but they are potentially a brake on harmful triumphalism.

For the rest, it seems important that people at large should be helped to a deeper understanding of what the scientific process is like. It's not a matter of education in the simple sense — knowing the structure of DNA, for example — but of understanding the necessarily tentative character of scientific conclusions, or theories, which all begin life as hypotheses.

One of the signs that distrust springs from people's differing perceptions of the world could be nicely illustrated by this experiment. It's a conceptual experiment, and it takes the form of a public opinion poll. It has to do with the way the human brain acquires its distinctive faculties. Two years ago, Francis Crick published a book with the title *The Astonishing Hypothesis*. The idea is that each human brain begins as a 'bag of neurons', or nerve cells, that spontaneously, in the heads of prenatal persons and young infants, form the interconnections that allow our brains to be the thinking machines they are. Presumably the hard-wired parts of the nervous system, the retina for example, form connections that are genetically determined, although it's only now that a start has been made on

the working of the genes involved in the embryonic development of the nervous system. What is still unclear is the degree to which the formation of the crucial interconnections between neurons is determined not only genetically but also by environmental and even random processes.

The experiment I recommend is simple: ask a group of scientists and a similar group of other people what they think of the proposal. My guess is that the scientists will not think Crick's hypothesis all that astonishing; in many ways it seems plain that this is indeed what happens. But the matching group of non-scientists will have a different reaction. Knowing that the brain is a physical synonym for 'mind', they will instinctively react against the hypothesis. Yet, when you think of it, this is the experimental ground on which the old question of the relative importance of nature and nurture will eventually be decided.

The research community, in its own interests, should be more ready than it is to acknowledge that issues like these are an essential part of the origin of distrust, that it should do what it can to anticipate them and that it should talk about them openly in the hope of making them go away. It simply amounts to a recognition that much of science seems to people at large to be a challenge to cherished beliefs. The need is not some kind of political correctness, as for example by pretending that quantum mechanics is not deterministic when it is, but for an open recognition by both sides that science does often challenge cherished beliefs, which are usually (but not always) religious beliefs.

If there is ever to be a *rapprochement* between science and those whose distrust springs from the fear that they will be made uncomfortable by whatever discoveries lie ahead, the chief responsibility for bringing it about must surely rest with those who are both scientists and religious people. By all accounts, there are many such people, many of whom who are prominent in public affairs.

The task should not be impossible. The record shows that the occasions in the past when a mechanistic explanation has been provided and accepted for phenomena previously regarded as somehow magical have not proved to be the death of what is called faith. The Stonehenge calendar is one example, the self-sustaining Expanding Universe has even been welcomed by religious people (because of the initial Big Bang), the origin of the elements (in stars like the Sun) has hardly caused a ripple in the world's pulpits and, although what is now known of the mechanism of life may eventually prove to be more subversive of contentment, there are still molecular biologists who go to church.



# Beating a path to Myc

Robert N. Eisenman and Jonathan A. Cooper

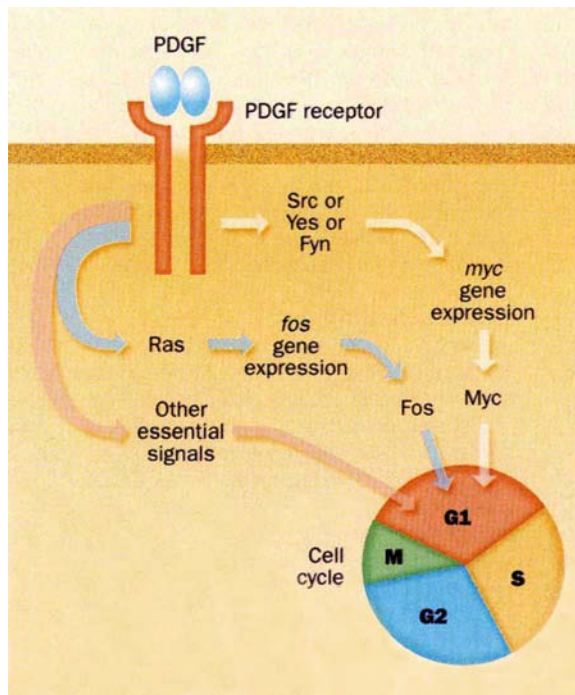
ARMCHAIR travellers of the world of signalling have become familiar with several ways in which extracellular growth signals reach the nucleus, stimulate gene expression and initiate changes in cellular behaviour. One such pathway, mediated by the small GTP-regulated protein Ras, has been used repeatedly during evolution to regulate gene expression and cytoplasmic events. In vertebrates, the Ras pathway is central for activating expression of certain genes that control the response to extracellular growth factors, but other pathways, such as that mediated by tyrosine kinases of the Jak family, are responsible for turning on other important genes. A newly discovered pathway by which growth factors regulate gene expression is reported by Barone and Courtneidge on page 509 of this issue<sup>1</sup>. The pathway makes use of the non-receptor tyrosine kinase Src, and bypasses Ras. The Ras pathway and the new mechanism are similar in that both are stimulated by the same growth factor receptor, both activate genes that encode transcription factors, and both contribute to the cascade of transcriptional responses that ultimately signal irreversible commitment to start DNA synthesis (S phase of the cell cycle) and to complete a cycle of cell division.

A well-known target of the Ras pathway is the *fos* gene, which encodes the transcription factor Fos and is induced rapidly when cultured mammalian fibroblasts are exposed to platelet-derived growth factor (PDGF)<sup>2</sup>. The target of the new pathway is the *c-myc* gene, which encodes the transcription factor, Myc, and is also induced by PDGF<sup>1</sup>. Myc, like Fos, has long been believed to be involved in rousing quiescent fibroblasts and driving progression through the cell cycle.

Both *fos* and *myc* are induced upon stimulation by mitogens or cytokines of diverse cell types, even though the ultimate response by the cell may be to differentiate into a specialized cell or to die off rather than to proliferate. This induced expression of both *fos* and *myc* messenger RNAs is considered a primary, or immediate early, response as it requires no new protein synthesis<sup>2,3</sup>. Compared to *fos*, however, *myc* mRNA induction is slow and *myc* RNA and protein expression continues throughout the cell cycle while

*fos* is rapidly downregulated<sup>4</sup>. Furthermore, there is ample evidence that *myc* gene expression is not only regulated by changes in the rate of transcriptional initiation, but also at the levels of transcriptional elongation, RNA processing or transport (reviewed in ref. 5).

Both classes of immediate early genes represented by *myc* and *fos* are important for the proliferative response of fibro-



The Src and Ras pathways, which are activated in response to binding of platelet-derived growth factor (PDGF) to its receptor on cell membranes. The activated dimerized PDGF receptor is associated with several proteins containing Src-homology 2 (SH2) domains, including the tyrosine kinases Src, Fyn and Yes. The Src-related tyrosine kinases link into the pathway proposed by Barone and Courtneidge<sup>1</sup>, which leads to activation of the transcription factor Myc. Other SH2-containing proteins link, via Ras, to activate expression of the transcription factor Fos. Together with other PDGF-dependent signals, Myc and Fos are necessary for the stimulation of resting fibroblasts to enter the cell cycle.

blasts. *myc* and *fos* are proto-oncogenes: constitutive expression of either can deregulate proliferation of fibroblasts, provided serum growth factors are present<sup>6</sup>. Under normal conditions, however, growth-factor-regulated expression of either gene alone is not sufficient for entry into the cell cycle, because blocking the expression of either *fos* or *myc* using antisense methods reduces or delays the start of S phase<sup>7,8</sup>.

A mutant form of the receptor for the colony-stimulating factor CSF-1 has been described which, while mediating normal expression of Fos and Jun, fails to induce Myc or promote progress

through the G1 phase of the cell cycle (the gap between the end of mitosis and the start of DNA replication)<sup>9</sup>; however, ectopic expression of *myc* restored the ability of the mutant receptors to induce cell proliferation. Thus the Myc and Fos transcription factors must regulate the expression of different genes required for transition through G1. Candidate targets for Myc include genes encoding enzymes required for DNA replication and possible activators of cyclin A and E expression<sup>10,11</sup>.

Although studies with mutant receptors suggested that separate signalling pathways for inducing *myc* and *fos* might exist<sup>9</sup>, the pathway from the PDGF receptor to Myc has remained largely uncharted. Courtneidge and her associates first showed that, upon activation, the PDGF receptor complexes with Src, and its close relatives Fyn and Yes, which are also tyrosine kinases<sup>12</sup>. The critical role of Src was established when injected antibodies against all three Src family proteins were found to completely inhibit PDGF-induced entry of NIH3T3 cells into S phase<sup>13</sup>. Src, Fyn and Yes normally bind to the tyrosine-phosphorylated PDGF receptor through their Src-homology 2 (SH2) domains, but the pathway can be blocked before it starts by dominant interfering Src or Fyn molecules that lack kinase activity, provided that they possess an intact SH2 domain. Once bound, Src is activated and can phosphorylate cell proteins that cannot be phosphorylated directly by the PDGF receptor.

Barone and Courtneidge<sup>1</sup> now show that an interfering form of Myc blocks entry into PDGF-dependent S phase, confirming that prevention of Myc function is sufficient to inhibit DNA synthesis. They then go on to show that induction of *myc* mRNA by PDGF is blocked by the Src-inhibitory antibodies. Fos is induced normally. The requirement for Src-family proteins in PDGF-induced entry into S phase can be circumvented by supplying Myc. By contrast, Fos and Jun do not bypass the block. The results add up to a model in which Src, Fyn and Yes are needed both for *myc* mRNA induction and S phase entry, and if these cytoplasmic tyrosine kinases are inhibited, *myc* expression becomes limiting for S phase entry — Src-family kinases are the toll house on the turnpike to Myc.

The model fits the predictions of the earlier studies where the mutant CSF-1 receptor that failed to induce *myc* was later found to activate Src poorly<sup>14</sup>. There is also a satisfying correspondence

between the prolonged requirement for Src kinase activity during G1 phase and the prolonged expression and requirement for Myc<sup>4,13,15</sup>. And there are indications that another cytoplasmic tyrosine kinase, Bcr-Abl, may require or induce myc expression through its SH2 domain<sup>16,17</sup>. It will be interesting to find out whether receptors that are not tyrosine kinases, such as cytokine receptors, induce Myc by the Src-dependent pathway.

Perhaps the most exciting follow-up will be the charting of the links between Src and myc mRNA induction. As the principal events that influence myc mRNA abundance during mitogenesis seem to occur at points other than transcriptional initiation, there may be signalling pathways that result in modulation of polymerase pausing or RNA turnover, both of which are crucial determinants of myc mRNA levels.

The abrogation of normal attenuation through pausing of myc transcription is thought to be a major factor in generating the increased amounts of Myc found in lymphomas<sup>5</sup>. It is conceivable that Src could indirectly or directly influence attenuation, for example by phosphorylating the newly identified von Hippel-Lindau tumour-suppressor protein (VHL) which promotes RNA polymerase II pausing<sup>18,19</sup>. Identifying the Src substrates involved in myc induction could explain why a receptor tyrosine kinase has to activate another tyrosine kinase in order to push a cell into S phase. □

Robert N. Eisenman and Jonathan A. Cooper are at the Fred Hutchinson Cancer Research Center, 1124 Columbia Street, Seattle, Washington 98104, USA.

## Overcoming inhibition

Clifford J. Woolf

THE consequences of a complete injury to the spinal cord are devastating — interruption of the normal flow of sensory input from the body to the brain and elimination of voluntary motor control below the break. Although there has been substantial progress in understanding growth in the central nervous system (CNS), the prospect of achieving clinically significant benefit, other than minimizing the immediate consequences of the injury, seemed very far off until the appearance of the study of Bregman and colleagues reported on page 498 of this issue<sup>1</sup>.

Dealing with a spinal-cord transection is immensely complicated. Imagine severing a conduit containing millions of cables that transfer information from highly ordered transmitters to specific receivers. Imagine, moreover, that cutting the cables leads to the permanent disablement of many of the transmitters; and that to prevent anomalous connections and maintain stability, the capacity for reconnection is actively inhibited. Finally try to appreciate that full function can only be restored if connections are re-established between specific transmitters and receivers.

Fortunately cautious optimism is now possible. Optimism, because by neutralizing a myelin-associated, neurite-growth-inhibitor molecule present on oligodendrocytes that repulses axonal growth *in vitro* and *in vivo*<sup>2</sup>, Bregman *et al.* have shown recovery of at least some motor function after injury to the adult mammalian spinal cord (their subjects were rats). Caution is necessary, however, because the recovery is partial and incomplete, the number of axons that have been induced to regenerate is very small, and the rat spinal cord is less dependent than the human on descending inputs to retain locomotor function after transection. Nevertheless this is a genuine step forward.

Caution is also necessary because the authors present no evidence that the corticospinal fibres induced to regenerate actually establish synaptic connections and are responsible for the functional recovery. Indeed, three types of change may all contribute to recovery after a partial spinal injury: compensation by the strengthening of existing connections; regeneration of injured axons; and collateral outgrowth of new axons from uninjured neurons. Which is the major contributor in the studies of Bregman *et al.* needs to be determined.

What are the factors that impede successful regeneration in the adult CNS? The first is failure of the injured neuron to survive. Axotomy, by depriving a cell of its normal trophic support from its

target, may result in atrophy and cell death. A therapeutic strategy here could be treatment with growth-factor supplements aimed at tyrosine kinase receptors expressed by injured cells. The second is the intrinsic growth capacity of the neuron. The assumption that neurons in the CNS of adult vertebrates cannot grow was shown by Aguayo and colleagues in the early 1980s to be incorrect; provision of a peripheral (as opposed to a CNS) nerve environment leads to some central axons growing long distances. But unfortunately many cells do not grow, and inducing them into a growing mode greatly enhances regeneration<sup>3</sup>. A permissive environment is necessary but not sufficient. What determines the growth capacity of neurons is not known, nor whether it is intrinsic or initiated by some extracellular signal. Potential clinical exploitation could involve providing a rate-limiting, growth-promoting protein. Transgenic mice engineered to constitutively overexpress one such protein, GAP-43, in adult neurons, show increases in both spontaneous and damage-induced axonal sprouting in the peripheral and central nervous systems<sup>4</sup>.

Finally we come to the potentially hostile environment. Here the issue is what aspects of cell-cell, cell-substrate or cell-growth-factor interactions are either permissive or repulsive for axonal growth. The list of candidates is long and constantly changing, and includes adhesion molecules, cell-surface and extracellular matrix molecules, proteoglycans, diffusible growth factors and myelin-associated inhibitory proteins and glycoproteins. Unresolved questions are whether there is retention of guidance cues laid down during development; the significance of changes in glial cells, the supporting cells of the CNS, on removal of contacts with axons; and the relative failure to remove axonal and myelin debris after CNS injury. What is clear, though, is that when seeded from cultures into the spinal cord, Schwann cells, the supporting cells of the peripheral nervous system, can promote axonal outgrowth<sup>5</sup>. Whether they do this by providing an obstacle-free trail or because they release growth and other factors is not known.

Bregman and colleagues' observation that neutralization of just one aspect of the environment, a myelin-associated neurite-inhibitory protein, is sufficient to promote some recovery of function is encouraging. Maybe there is a threshold for producing regeneration — perhaps not all obstacles to regrowth have to be removed. Further progress is contingent on the purification and cloning of this

1. Barone, M. V. & Courtneidge, S. A. *Nature* **378**, 509–512 (1995).
2. Greenberg, M. E. & Ziff, E. B. *Nature* **311**, 433–438 (1984).
3. Kelly, K., Cochran, B. H., Stiles, C. D. & Leder, P. *Cell* **35**, 603–610 (1983).
4. Hann, S. R., Thompson, C. B. & Eisenman, R. N. *Nature* **314**, 366–369 (1985).
5. Spencer, C. A. & Groudine, M. *Adv. Cancer Res.* **56**, 1–48 (1991).
6. Evan, G. I. *et al.* *Cell* **69**, 1–20 (1992).
7. Heikkilä, R. *et al.* *Nature* **328**, 445–449 (1987).
8. Nishikura, K. & Murray, J. M. *Molec. cell. Biol.* **7**, 639–649 (1987).
9. Roussel, M. F., Cleveland, J. L., Shurtleff, S. A. & Sherr, C. J. *Nature* **353**, 361–363 (1991).
10. Bello-Fernandez, C., Packham, G. & Cleveland, J. L. *Proc. natn. Acad. Sci. U.S.A.* **90**, 7804–7808 (1993).
11. Jansen-Durr, P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **90**, 3685–3689 (1993).
12. Kypta, R. M., Goldberg, Y., Ullg, E. T. & Courtneidge, S. A. *Cell* **62**, 481–492 (1990).
13. Twamley-Stein, G., Pepperkok, R., Ansorge, W. & Courtneidge, S. A. *Proc. natn. Acad. Sci. U.S.A.* **90**, 7696–7700 (1993).
14. Courtneidge, S. A. *et al.* *EMBO J.* **12**, 943–950 (1993).
15. Waters, C. M., Littlewood, T. D., Hancock, D. C., Moore, J. P. & Evan, G. I. *Oncogene* **6**, 797–805 (1991).
16. Cleveland, J. L., Dean, M., Rosenberg, N., Wang, J. Y. J. & Rapp, U. R. *Molec. cell. Biol.* **9**, 5685–5696 (1989).
17. Afar, D. E. H., Goga, A., McLaughlin, J., Witte, O. N. & Sawyers, C. L. *Science* **264**, 424–426 (1994).
18. Duan, D. R. *et al.* *Science* **269**, 1402–1406 (1995).
19. Kibel, A., Iliopoulos, O., DeCaprio, J. A. & Kaelin, W. G. *Science* **269**, 1444–1446 (1995).

protein, and for instance on finding out whether it belongs to or resembles the semaphorin families of growth-cone-collapsing proteins which have a critical role in directing developing neurons to appropriate targets<sup>6</sup>. Subsequent issues will be that of the best way to neutralize its activity in patients, and whether human recombinant versions of the neutralizing IN-1 monoclonal antibody used by Bregman *et al.* can be engineered (and, if they can, whether they will be able to get to the inhibitory protein along deep fibre tracts in the human spinal cord below a lesion). An alternative approach may well be to modify the reaction of the growth cone to this inhibitory signal by interfering with receptors or signal transduction mechanisms reducing the effects of inhibitory

or repulsive signals<sup>6</sup>. We also need to understand why embryonic neurons implanted into the adult CNS find myelinated fibre tracts permissive for axonal growth<sup>7</sup>. Is it because they do not express the receptor for the myelin-associated neurite inhibitor?

Many promising research avenues beckon. But again, given that so much is at stake for patients with spinal injury, if one must err it should be on the side of caution. No effective therapy is in immediate sight. And it should be remembered that not all axonal regeneration is functionally adaptive; that a rewiring of spinal circuitry resulting from the growth of sensory axons after nerve injury in the adult may contribute to chronic pain<sup>8</sup>; and that it is quite possible that the encouragement

of growth may not always recapitulate the connections made during development, with consequent functional problems. □

Clifford J. Woolf is in the Department of Anatomy and Developmental Biology, University College London, Gower Street, London WC1E 6BT, UK.

1. Bregman, B. S. *et al.* *Nature* **378**, 498–501 (1995).
2. Schnell, L. & Schwab, M. E. *Nature* **343**, 269–272 (1990).
3. Richardson, P. M. & Issa, V. M. K. *Nature* **309**, 791–793 (1984).
4. Aigner, L. *et al.* *Cell* **83**, 269–278 (1995).
5. Xu, X. M., Guenard, V., Kleitman, N. & Bunge, M. B. *J. comp. Neurol.* **351**, 145–160 (1995).
6. Puschel, A. W., Adams, R. H. & Betz, H. *Neuron* **14**, 941–948 (1995).
7. Davies, S. J., Field, P. M. & Raisman, G. J. *Neurosci.* **14**, 1596–1612 (1994).
8. Woolf, C. J., Shorthand, P. & Coggeshall, R. E. *Nature* **355**, 75–77 (1992).

## OBITUARY

## Gérard de Vaucouleurs (1918–95)

GÉRARD de Vaucouleurs, who died in Austin, Texas, on 7 October 1995, came into extragalactic astronomy 45 years ago from outside the mainstream, at a time when there was little interest in understanding galaxies and the physics of galaxies. He was no theorist, but he knew what galaxies looked like and how they appear to be distributed. Modern astronomy needs more like him. By now we have moved to the other extreme, a situation in which we have many theorists and others trying to build models and understand galaxies who have never looked at the optical or radio images in any detail, and thus are unfamiliar with what they really need to explain.

De Vaucouleurs, who took the maiden name of his mother, was born in Paris and educated at the Sorbonne. In 1950 he came to England and worked for a period for the French Section of the BBC in London. I remember his enthusiasm and his dedication to astronomy at that time, and these characteristics were to endure.

In his early career de Vaucouleurs was extensively involved in popular astronomy, astronomical photography and science writing. His main interests professionally were in the planet Mars, and in the extragalactic Universe. Very early on he wrote an extensive monograph on the planet Mars. In 50 years he published about 400 papers, and many books and popular articles. But the contribution for which he will be remembered was to the study of galaxies. In the 1950s, when de Vaucouleurs began his work, a large part of astrophysics was concerned with the structure and evolution of the stars, and comparatively few astronomers were working on galaxies. This is a great contrast to the situation today

when extragalactic astronomy and cosmology are some of the most active areas of research, with hundreds of investigators.

Above all, de Vaucouleurs was a collector of data, with an extraordinary knowledge of the galaxies. He knew the galaxies — he classified thousands of them, and as a cataloguer he was with-

cases new photometric observations. De Vaucouleurs revised and refined the original Hubble classification of galaxies by introducing a three-dimensional scheme which was valuable, if perhaps a little too complicated at times. He also worked on the distance scale of the Universe and the determination of the Hubble constant. In the 1970s and 1980s he argued for a large value for the Hubble constant, around  $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , as compared with the small value of about  $50 \text{ km s}^{-1} \text{ Mpc}^{-1}$  advocated by Sandage and Tamman. This debate is still going on.

He worked extensively on the Magellanic Clouds and on the Local Group of nearby galaxies. In 1959 with Frank Kerr he obtained the first reliable mass estimates of the Magellanic Clouds. He carried out detailed surface photometry on galaxies and was particularly well known for his work on the luminosity distribution and structure of elliptical galaxies.

Late in his career, de Vaucouleurs' abilities and his work were recognized by his peers. One of his most important feats was to persuade the community of the reality of the local supercluster of galaxies with the Virgo cluster as its nucleus. This concept of the local supercluster, and its extension to other large superclusters, was strongly resisted for many years by other established extragalactic astronomers, but has now become widely accepted.

In appearance he was small and always impeccably dressed; after he moved to Texas he often dressed accordingly. He is survived by his second wife Elysabeth. Geoffrey Burbidge

Geoffrey Burbidge is in the Center for Astrophysics and Space Sciences, University of California, San Diego, La Jolla, California 92093, USA.

IMAGE  
UNAMABLE  
FOR COPY RIGHT  
FOR COPY RIGHT  
REASONS

Dept. of Astronomy, Univ. Texas

out peer, although, as he would freely admit, he was not an astrophysicist.

Probably his most enduring contribution lay in providing huge amounts of data for three *Reference Catalogues of Bright Galaxies* published in 1964, 1976 and 1991. This, like much of his other work, was carried out jointly with his first wife, Antoinette, who died in 1987, and with others at the University of Texas. Production of these catalogues involved a prodigious amount of data handling, analysis and in many



# A very cool customer

France Allard

Do brown dwarfs — stars that cool and fade without ever achieving stable hydrogen burning — really exist? Over the past year, concerted hunts have flushed out a handful of likely brown dwarf candidates hiding in the solar neighbourhood<sup>1,2</sup>, and on page 463 of this issue, Nakajima *et al.*<sup>3</sup> report their finding of the coolest and lowest-mass candidate yet.

Star formation theory gives no reason why brown dwarfs should not exist. Proto-stellar clouds should be able to fragment into objects only 12 to 74 times the mass of Jupiter, massive enough to trigger nuclear deuterium fusion but unable to sustain the stable hydrogen burning which powers stars for billions of years. Extrapolation of the observed mass distribution of stars suggests that there should be myriads of such 'failed' stars, more than  $10^5$  brown dwarfs for every Sun-like star we see. The indications of dark matter throughout the Universe seem to go hand in hand with this postulated population of objects which would fade into darkness after their original deuterium reserves were exhausted in a scant 10 million years. So where are they?

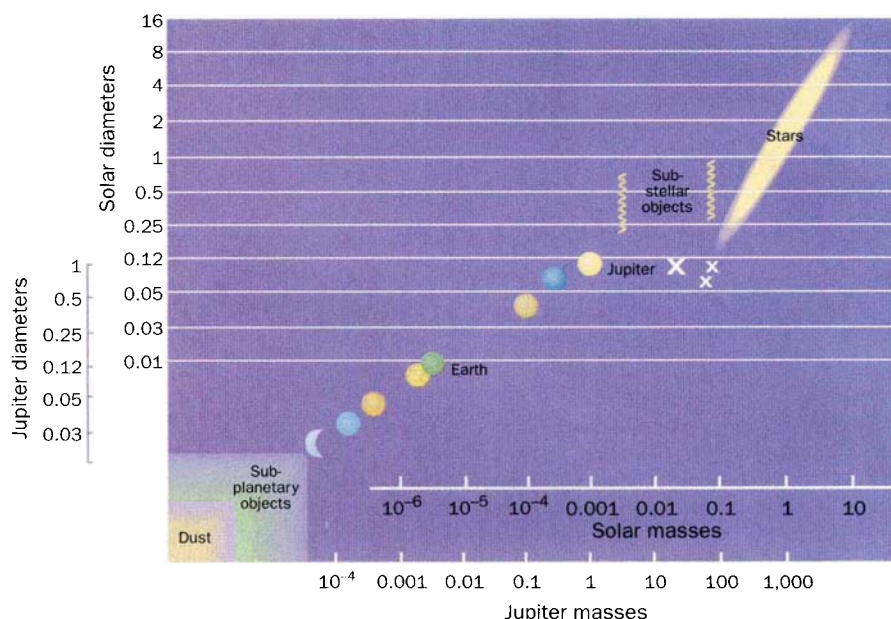
Admittedly, they are elusive prey. When a brown dwarf is young and still burning deuterium, it can be as bright as one of its stellar cousins and as easily detected. Unfortunately, it is also difficult to distinguish a young brown dwarf from a normal cool star unless it is in a binary system whose orbit yields a direct mass estimate. Later in life, the same brown dwarf may be too faint to detect at all. Despite such problems, the past year has seen a smattering of likely detections: Rebolo *et al.*<sup>1</sup> reported the discovery of brown dwarfs Teide 1 and 2, and Stauffer *et al.*<sup>2</sup> that of PPI 15 in the Pleiades star cluster. Of the confirmed or suspected brown dwarfs, all have masses estimated to lie close to the stellar limit of about 0.08 solar masses. (The estimated luminosity of Teide 1 suggests a mass as low as 0.02 solar masses, but this is not supported by evidence for the cool atmosphere expected for a very-low-mass brown dwarf.) The next most massive substellar object available for study is a mere 0.001 solar masses: the planet Jupiter in our own Solar System.

The discovery by Nakajima *et al.* should help to fill the gulf between jovian planets and low-mass stars. After a painstaking search for substellar objects around nearby stars, they have identified an object, named Gliese 229B after its brighter companion star, which is the coolest and lowest-mass brown dwarf known, weighing in at about 20 Jupiter masses (0.02 solar masses). In many ways,

it is more reminiscent of a planet than our conventional picture of a star.

The keys to recognizing such a substellar object are in its spectrum. The earliest generations of stars are believed to start with a primordial supply of lithium left over from Big Bang nucleosynthesis. Even the coolest main-sequence stars are hot enough to burn lithium in their cores, and thanks to convection, all the lithium is eventually brought to the interior and

by Tsuji and colleagues<sup>5</sup> show that methane features begin to appear in the infrared. At roughly 1,500 K, titanium dioxide and silicate clouds form at the expense of TiO, profoundly modifying the thermal opacity of the atmosphere. Below 1,000 K, methane and additional water are able to form at the expense of carbon monoxide. In models with temperatures below 200 K, molecular nitrogen can combine with the abundant hydrogen to become gaseous ammonia, and additional water clouds condense at a depth where the pressure is about 1 bar. The high ammonia clouds combined with the water clouds modify the reflective proper-



Gulf in mass between the heaviest planet (Jupiter) and lightest stars. Previous masses of brown dwarf candidates cluster near the small crosses. The new, cool brown dwarf<sup>3</sup> (large X, diameter uncertain) at about 20 Jupiter masses, should help to bridge the gap.

destroyed. Brown dwarfs below about 0.06 solar masses are not hot enough long enough to deplete their lithium supply. Therefore, the presence of lithium in the spectrum of a low-mass star is a definite (but not necessary) signature of a brown dwarf. At the time of writing, only one brown dwarf candidate, PPI 15, has passed this test, based on high-resolution spectra obtained with the Keck telescope by Basri and co-workers<sup>4</sup>; the other suspects await similar confirmation.

Less-massive brown dwarfs are expected to have even more distinctive spectral signatures, thanks to the unique changes in molecular chemistry that occur across the atmospheric temperature range from 2,000 K (the coolest stars) to 150 K (jovian planetary conditions). At 2,000 K, most of the carbon is locked into carbon monoxide, whereas most of the oxygen is found in titanium and vanadium monoxides (dominating the optical absorption) and water vapour (shaping the infrared spectrum). As the effective temperature drops below 1,800 K, models

ties of the atmosphere at these jovian temperatures.

On these grounds, Gliese 229B is a very cool little object indeed. It is the first extrasolar object ever found to show the presence of methane in its spectrum. Its strong infrared methane and water bands suggest an effective temperature well below 1,800 K, and perhaps as low as 900 K. Rather than a small failed star, could Gliese 229B be a giant planet? Until its mass can be derived dynamically, it must be estimated from theoretical arguments. Extinguished brown dwarfs and giant planets shine naturally by release of residual internal heat from gravitational contraction and isotopic decay. So the more massive the object, the brighter it is at any time. The rate of cooling with time can be modelled to provide a mass estimate for the object on the basis of its luminosity and age. Nakajima *et al.* find Gliese 229B to have a luminosity of  $2 \times 10^{-6}$  times that of the Sun. According to recent models for extrasolar giant planets by Saumon *et al.*<sup>6</sup>, a 12-Jupiter-mass planet as old as

Gleise 229A (the normal M dwarf star in the system) would be about 30 per cent fainter than Gleise 229B with an effective temperature of only 500 K. But an object as cool as Gleise 229B emits a significant fraction of its energy at wavelengths far beyond 2  $\mu\text{m}$ ; the estimated luminosity is likely to increase as more of its hidden infrared tail is revealed. A remote possibility remains that Gleise 229B is a foreground object which is aligned by chance with the M dwarf, leading to a lower luminosity and making it more likely to be a planet. But it would then be an orphaned planet, without a visible parent star.

Although the exact mass of Gleise 229B is still in doubt, there seems little question that it lies comfortably below the hydrogen-burning mass limit and bridges the gap between low-mass stars and Jupiter (see figure). The former are believed to form by the collapse of interstellar cloud fragments, the latter by

accretion within a circumstellar disk. What about the birth of a transitional object like Gleise 229B? This is a question theorists must now face in testing current models of both star and planet formation. Brown dwarfs may be too few to account for the matter which remains unseen in our Galaxy, but the handful discovered so far could shed new light on how baryonic matter was able to coalesce into the stars and planets we can see.  $\square$

France Allard is in the Department of Physics, Wichita State University, Wichita, Kansas 67260, USA.

1. Rebolo, R., Zapatero Osorio, M. R. & Martin, E. L. *Nature* **377**, 129–131 (1995).
2. Stauffer, J. et al. *Astr. J.* **108**, 155–159 (1994).
3. Nakajima, T. et al. *Nature* **378**, 463–465 (1995).
4. Basri, G., Marcy, G. & Graham, J. R. *Astrophys. J.* (in the press).
5. Tsuji, T., Ohnaka, K. & Aoki, W. in *The Bottom of the Main Sequence — and Beyond* (ed. Tinney, C. G.) 45–49 (Springer, Berlin, 1995).
6. Saumon, D. et al. *Astrophys. J.* (in the press).

## ECOLOGY

# Location is a sticky business

Peter D. Moore

It is said that there is no such thing as a free lunch, and that maxim certainly applies to insectivorous plants, for they need to allocate much of their energy to enticing and trapping prey. The costs may be even greater if the plant has to grow under suboptimal conditions simply to ensure that its lunch will come stumbling by, and this is a problem that Regino Zamora examines in a paper<sup>1</sup> describing the ecological quandary facing the scarce butterwort, *Pinguicula vallisneriifolia*, which is found only in southern Spain. Should a butterwort sit in the sun and tap solar energy, or skulk in the shade where the insects lurk? The energetic and nutrient equations are not easy to balance.

## Glands

Plants that have adopted the predatory habit can only sit and wait for their prey, but some have developed pigments that appeal to insects, as in the case of some pitcher plants such as *Sarracenia*, which occupy open, exposed locations where their presence attracts the attention of flying insects. Other insectivorous plants, however, hunt on a more random basis, their success depending upon being located in a position where their prey is likely to wander. The butterworts are generally of this type, having leaves that lack attractive pigments and operate on a 'sticky-trap' principle. Scanning electron micrographs of *Pinguicula* leaves show<sup>2</sup> that they carry two types of glands, one sessile and the other stalked. The stalked

glands seem to operate in the initial stages of trapping and the sessile glands contribute to the digestive process<sup>3</sup>.

Typical butterworts have broad, flat, sticky leaves forming a basal rosette from which the flowering stems arise. *Pinguicula vallisneriifolia* is unusual in that it grows on the vertical faces of wet limestone cliffs. Its first leaves form a rosette, but later leaves extend outwards from the cliff face and then droop downwards, sometimes extending for 30 cm. Such sticky, hanging structures are effective in trapping insects living close to the cliff face, those leaves that extend out from the wall being considerably more efficient than the original basal rosette.

Zamora set out to determine the optimal habitat for prey capture and selected four sites that varied in their degree of shading. At each site, and for each of six sampling occasions in summer, selected leaves were stripped of all prey and the arrival of new victims was recorded periodically over a 48-hour period. All of the fresh prey material was then removed, identified and its biomass estimated. At the same time, artificial sticky traps of similar size and form to the butterwort leaves were placed in each of the four habitats to provide an indication of prey availability. Allowance had to be made, of course, for the fact that larger insects (greater than 5 mm) were not effectively trapped by the stalked glands of the *Pinguicula* leaves, so only the smaller arthropods were considered.

A further complication is that a collec-

tion of stuck-up bugs offers a tempting opportunity for other, more mobile, insectivores to obtain that proverbial free lunch, and animals of such low moral fibre were present in both sunny and shady habitats in the form of lizards in the sunny location and kleptoparasitic slugs in the shade. No doubt the long, deadly, drooping leaves of *P. vallisneriifolia* proved too difficult a prospect for smaller animals such as ants that have been recorded as important leaf-robbers in other *Pinguicula* species<sup>4,5</sup>. Fortunately for the experimenter, however, the rates of theft were low and were roughly equivalent in the different habitats, so did not interfere with the overall analysis.

## Viscosity

Zamora found that the highest concentrations of insects (as determined by the success of the artificial sticky traps) occurred in the shady, moist habitats. Similarly, the greatest insect-trapping success rates were also recorded in the shade. Insects tended to avoid the warm, dry sunny locations and rates of prey capture were poorer in such situations. The retention of trapped animals, however, proved more effective in the sunny habitats because the greater evaporation led to the development of a more viscous secretion from the leaf glands, so this may in part compensate for the poorer trapping success. Low viscosity means that the bigger, stronger prey escape, which tends to occur more in the shady locations. So, when measured in terms of biomass, the plants in intermediate shade were actually more successful in their trapping activities than either those in the sun or in the dense shade.

*Pinguicula vallisneriifolia*, like other insectivorous plants, grows in locations where certain plant nutrients are in poor supply, hence its need to collect nitrogen and phosphorus from animal tissues. Its optimum location for overall growth, therefore, will involve a trade-off between its need for light as a photosynthetic resource and its limitation by nitrogen or phosphorus (or both), demanding a shady life-style. Perhaps the *Pinguicula* species of higher latitudes have the advantage here, for insects in cooler climes often congregate in the sunnier spots to bask, and this means that on a Scottish mire a butterwort can stretch out in the sun and collect the best of the bugs at the same time.  $\square$

Peter D. Moore is in the Division of Life Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

1. Zamora, R. *Oikos* **73**, 309–322 (1995).
2. Heslop-Harrison, Y. & Heslop-Harrison, J. *Ann. Bot.* **47**, 293–319 (1981).
3. Heslop-Harrison, Y. & Knox, R. B. *Planta* **96**, 183–211 (1971).
4. Zamora, R. *Oikos* **59**, 368–372 (1990).
5. Moore, P. D. *Nature* **350**, 192 (1991).



# Characteristic or haphazard?

Ross S. Stein

THE next time you gaze out of your window, only to find that the nearby lake has been drained and an excavation suitable for the foundations of the World Trade Center has been dug, don't call the city council. It's probably just a group of geologists dating prehistoric earthquakes on a fault passing underfoot. In their search to find sites with a longer and more precise history of past earthquakes, palaeoseismologists are being enticed underwater by the finely layered, carbon-rich strata of lakes and ponds, and they are digging not simple slices across faults but a deep labyrinth of excavations, tracing offset stream channels across fractures large and small. And they are benefiting from new on-site or 2-day-turnaround radiocarbon dating so they know at the outset if they have hit paydirt. Why bother? Because, as revealed at a recent meeting\*, it is our best hope of finding out if large earthquakes are truly characteristic: in other words, whether events of the same size, or at least the same slip, recur on a given fault or fault segment<sup>1-3</sup>.

Speaking like a rabbi whose role is not only to comfort the afflicted but also to afflict the comfortable, R. Muir-Wood (EQE, Gloucester) argued that given what is now known about fractal, chaotic and self-organized critical systems, it should be surprising to find any regularity in earthquake occurrence. Have we become too enamoured — too comfortable — with the utility of characteristic earthquakes, and the associated repeat time for such events? Despite an eloquent defence for their existence in the prehistoric record (D. Schwartz, US Geological Survey, Menlo Park; K. Sieh, California Inst. Technol.; T. Rockwell, San Diego State Univ.), there is no avalanche of evidence for a concept deeply embedded in probabilistic assessments of earthquake occurrence. Most convincing are trench excavations along some of California and Utah's major faults, where the displacement at a site is similar for several past earthquakes. The slip in the 1983 magnitude 7 earthquake on the Lost River fault in Idaho also looks identical to an event about 7,000 years ago. But few trenches uncloak more than two or three earthquakes, and flights of marine terraces in New Zealand, which record the

sudden uplift associated with offshore earthquakes, do not paint a picture of similar-sized events<sup>4</sup>. What is needed is a much longer record of earthquake slip: hence the deeper excavations.

Whether the largest earthquakes on a fault recur with the same size is clouded by another phenomenon so widely reported at the gathering that it came close to a meeting mantra: clustering. In Iran, Italy,

lier episode? Excavations also leave open the possibility that 500 km of the southern San Andreas fault ruptured in about AD 1480<sup>5</sup>, which would yield an event of magnitude 8.1, 180 km longer than the magnitude 7.9 event in 1857 currently used to infer a maximum magnitude. But a cluster of three magnitude 7.6 events over 50 years satisfy the data as well, much as 725 km of the North Anatolian fault in Turkey ruptured in four large earthquakes during the Second World War. Thus assigning a maximum earthquake magnitude — and hence the maximum ground shaking — to a fault becomes harder to estimate from the prehistoric record if clustering is common. More precise age-dating is the only way out: hence the drained lakes.

Ironically, exciting progress has been achieved in tracking down the most concealed and elusive of earthquake sources: blind thrust and normal events in which the fault does not appear at the ground surface. Taking advantage of a lake repeatedly filled and drained by nature, M. Meghraoui (Univ. Cergy-Pontoise) dated seven prehistoric earthquakes at the site of the magnitude 7.3 shock at El Asnam, Algeria, from the damming of a river by co-seismic fold uplift<sup>6</sup>. At the Istituto Nazionale di Geofisica (Rome), researchers have carried out a systematic inventory of normal faults tracing the spine of the southern Apennines that are partially concealed by a former episode of compression, finding historical and prehistoric events separated by 1,000–3,000 years (see figure)<sup>9,10</sup>. Drawing from Iran's long and tragic historical records of earthquakes, M. Berberian (Najarian Assoc., New Jersey) identified numerous previously unknown blind thrust earthquake sources in the Zagros fold belt<sup>11</sup>. The 12 million people now living in Tehran are exposed to an earthquake potential in great need of study.

But such use of rich historical records



Fault segmentation, slip rate ( $S_f$ ) and earthquake recurrence time ( $T_r$ ) along the central and southern Apennine chain, Italy. Faults known from field observations or inferred from instrumental data are in red. Faults inferred from geological or historical data are in purple (a, high confidence level; b, low confidence). Shading marks weak zones with moderate seismogenic potential. Dates, where given, are for the last known event. (Courtesy of the Istituto Nazionale di Geofisica, Rome.)

Greece, Turkey, New Zealand, China, California and Japan, historical earthquakes cluster in space and time, consistent with studies of instrumentally recorded shocks<sup>5</sup>. Rather than behaving as isolated events, earthquakes must thus interact. Mounting evidence points to one shock triggering the next by the transfer of stress<sup>6,7</sup>. But rarely is it possible to distinguish from trench logs whether near-synchronous displacements along a fault represent one large earthquake or a cluster of smaller ones spanning several decades. A series of trenches has revealed, for example, that nearly the entire fault that ruptured in the 1992 magnitude 7.4 earthquake in Landers, California, slipped about 5,500 years ago. Is the 1992 event therefore characteristic, or could a cluster of smaller shocks have struck along the fault during the ear-

\*Active Faulting Studies for Earthquake Hazard Assessment, Ettore Majorana Centre for Scientific Culture, Erice, Sicily, 27 September–5 October 1995. Proceedings available by email from <Erice95@8800.ingrm.it>.

- Schwartz, D. P. & Coppersmith, K. J. *J. geophys. Res.* **89**, 3051–3058 (1984).
- Bakun, W. H. & McEvilly, T. V. *J. geophys. Res.* **89**, 3051–3058 (1984).
- Sieh, K. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Ota, Y., Hull, A. G. & Berryman, K. R. *Quat. Res.* **35**, 331–346 (1991).
- Kagan, Y. Y. & Jackson, D. D. *Geophys. J. int.* **104**, 117–133 (1991).
- Stein, R. S., King, G. C. P. & Lin, J. *Science* **265**, 1432–1435 (1994).
- Harris, R. A. et al. *Nature* **375**, 221–224 (1995).
- Meghraoui, M. & Doumaz, F. *J. geophys. Res.* (in the press).
- Valensise, G. & Pantosti, D. *Terra Nova* **4**, 472–483 (1992).
- Pantosti, D., Schwartz, D. P. & Valensise, G. *J. geophys. Res.* **98**, 6561–6577 (1993).
- Berberian, M. *Tectonophysics* **241**, 193–224 (1995).
- Ambraseys, N. N. & Finkel, C. F. *The Seismicity of Turkey and Adjacent Areas: A Historical Review, 1500–1800* (Muhittin Salih EREN, Istanbul, 1995).

to elucidate earthquake occurrence<sup>12</sup> is, sadly and perhaps shamefully, rare. Muir-Wood and D. Giardini (Univ. Roma) remonstrated with the profession for devoting so few resources to seeking and husbanding our trove of historical earthquakes. These documents bear strongly on the issues of characteristic earthquakes, clustering and blind events, and

because teasing science from diaries and church archives demands skills that few today possess, the historical treasure may be far more perishable than the strata in a trench. □

Ross S. Stein is with the US Geological Survey, 345 Middlefield Road, Menlo Park, California 94025, USA.

## NONLINEAR DYNAMICS

# Ordering chaos with disorder

Steven H. Strogatz

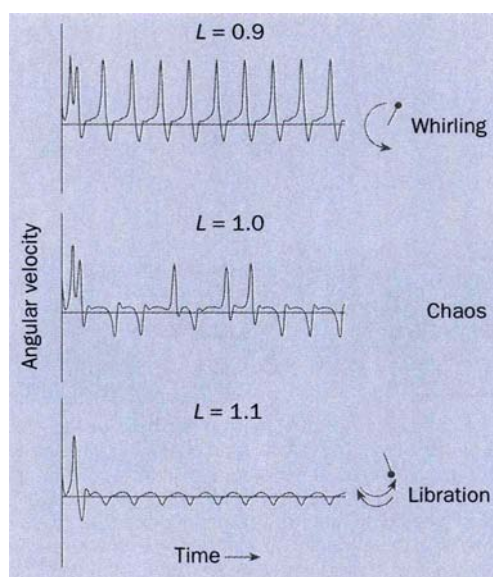
SYNCHRONIZATION of an array of nonlinear oscillators — for instance, coupled Josephson junctions or semiconductor lasers — is most easily achieved if all the oscillators are identical. Or so one might have thought. On page 465 of this issue, Braiman, Lindner and Ditto provide an intriguing counterexample. They start with an array of periodically forced pendula that lapses into spatiotemporal chaos if the pendula are identical. Yet when disorder is introduced into the array, the system snaps into periodic behaviour. In a sense, chaos has been tamed by disorder.

Oscillator arrays and spatiotemporal chaos are both relatively recent themes in nonlinear dynamics. In the late 1970s and throughout the 1980s, theorists focused on systems with few degrees of freedom. This was the heyday of the single damped driven pendulum, the Lorenz equations, and other small systems of ordinary differential equations, along with the logistic map, the Hénon map, and other simple iterated mappings. The key discovery from this era, due to Feigenbaum and others, was that there are universal laws governing the transition from periodic to chaotic behaviour: roughly speaking, diverse physical systems can go chaotic in the same way. This universality was beautifully explained by renormalization group ideas borrowed from statistical mechanics, and was later confirmed in experiments on fluids, electronic circuits, oscillating chemical reactions, semiconductors and mechanical systems (see, for example, P. Cvitanovic (ed.) *Universality in Chaos* 2nd edn; Hilger, Bristol, 1989).

Some of these successes seemed too good to be true. How could a spatially extended system, such as a fluid undergoing thermal convection, be governed by laws derived from simple mappings that contained none of the physics? The catch is that all the successful experimental tests were done on highly dissipative, spatially constrained systems. By design, these systems acted as though they had only a few

degrees of freedom, so they naturally obeyed universality theory.

The outstanding problem today (and it will be with us for decades to come) is to develop techniques for understanding systems with many degrees of freedom, such as turbulent fluids or fibrillating hearts. As a first step, one approach is to investigate idealized caricatures of such complex



The motion of a single periodically driven pendulum depends on its length  $L$ . Both short and long pendula settle into periodic motions, either whirling (top panel) or libration (bottom panel), whereas a pendulum of intermediate length displays a chaotic sequence of whirling and libration.

systems. It is in this spirit that Braiman *et al.* have turned their attention to arrays of nonlinear oscillators.

Each of their oscillators is an underdamped pendulum driven periodically by a combined d.c. and a.c. torque (the net torque is always positive, but it varies sinusoidally). In response to such a drive, a single pendulum exhibits three types of long-term motion, depending on its length (see figure). For the parameters used by Braiman *et al.*, a pendulum of dimensionless length 0.9 rapidly settles into a periodic whirling motion, completing exactly one full rotation for each cycle of the dri-

ving torque. Frequency-locking to the a.c. part of the drive also occurs if the pendulum has length 1.1, but now the pendulum hangs downwards at all times, librating once about the downward vertical per drive cycle. For intermediate length 1.0, the pendulum's motion is chaotic; irregular librations are sporadically interrupted by rotations over the top.

Braiman *et al.* consider a chain of such pendula coupled to their nearest neighbours by torsional springs. In a homogeneous chain where all the pendula have length 1.0 (as in Fig. 1a of Braiman *et al.*), the chaos of the individual oscillators persists in the coupled array; all the pendula whirl and librate, but with no repeating spatial or temporal structure. However, if the lengths of the pendula are randomly distributed across the chain, say with a 10 or 20 per cent uniform spread, the array breaks up into well defined domains that are frequency-locked (at various ratios) to the a.c. drive. An explanation given by Braiman *et al.* seems entirely plausible, in light of our discussion — the disorder creates subpopulations of long and short pendula that can easily frequency-lock to the a.c. drive. It does not matter whether these locked pendula are librating or whirling; they impose their rhythm, via the coupling, on those that would otherwise be chaotic, thus taming the chaos in the entire array.

A much more dramatic effect would be to tame chaos through disorder while keeping all the pendula in their chaotic regimes. Braiman *et al.* claim that they have observed this too, in which case a more subtle mechanism must also be at play.

Future work should test whether disorder-induced taming is common in spatially extended systems, and quantify the likelihood of its occurrence. Regarding potential applications, Braiman *et al.* argue that disorder might be used to 'control' spatiotemporal chaos, but that seems optimistic, given that no one knows what complex periodic pattern will arise after disorder is introduced. Nevertheless, taming might still be useful in mode-locking applications where any kind of periodic behaviour is preferred to chaos. For instance, superconducting Josephson arrays or semiconductor laser arrays are two systems in which collective locking is technologically desirable, but notoriously difficult to achieve. Normally the difficulty is blamed on the inevitable inhomogeneities, but now we have to wonder whether better results might be achieved by disordering the systems even further. Admittedly, the result will not be coherent oscillation, but perhaps a complex spatial pattern of frequency-locking is better than no locking at all. □

Steven H. Strogatz is in the Department of Theoretical and Applied Mechanics and the Center for Applied Mathematics, Cornell University, Ithaca, New York 14853, USA.



# Strong case for weak bosons

Dieter Zeppenfeld

A DOZEN years after the discovery of the intermediate vector bosons  $W^+$ ,  $W^-$  and  $Z$  at CERN, the CDF and D0 collaborations at Fermilab in the United States have observed the production of pairs of vector bosons. Various combinations have been searched for, and the two collaborations now report, in a cluster of papers in *Physical Review Letters*, that they have observed the production of weak bosons associated with photons,  $\gamma$  ( $W\gamma$  production<sup>1,2</sup> and  $Z\gamma$  production<sup>3,4</sup>) and a probable  $W^+W^-$  pair production event<sup>5</sup>, together with much improved upper bounds on the production rate of  $WW$  and  $WZ$  pairs<sup>6</sup>. How do these findings fit into the overall picture of forces between elementary particles?

In the proton-antiproton collisions which are being studied at Fermilab, single weak bosons as well as vector boson pairs are generated in the annihilation of quarks and antiquarks, the constituents of the colliding protons. There are two different ways in which a quark and an antiquark may annihilate to produce two vector bosons (see box). One possibility is that the vector bosons are radiated directly from the incident particles. This process, which is depicted in the first part of the figure, is also responsible for the annihilation of an electron and a positron into two photons. A second possibility, which is forbidden for two photons, is the annihilation of the quark and antiquark into a virtual boson which then splits into the pair of vector bosons which is observed in the experiment.

This second contribution depends on the way in which the vector bosons interact with each other. Via the pair production process, it now becomes possible to study the simultaneous interactions of three vector bosons in a systematic way. It is the direct measurement of these three-boson vertices, as they are known, which is at the heart of the present excitement. It allows a first direct test of whether the  $W$ , the  $Z$  and the photon are indeed the gauge bosons of a non-abelian gauge symmetry, as stipulated in the standard model of particle physics.

Within the gauge theory framework the  $WW\gamma$  and  $WWZ$  vertices are fixed precisely. For example, the magnetic moment of the  $W$ ,  $g_w$ , is predicted to be two 'W magnetons', in complete analogy to Dirac's famous result for the electron. Similarly, quantities like the electric dipole and quadrupole moments of the  $W$  and analogous quantities for the  $WWZ$  vertex are uniquely fixed by the gauge symmetry of the standard model. Exactly for these gauge theory values the two contributions in the figure show strong

destructive interference: there are 'gauge theory cancellations' between the two contributions and these cancellations lead to cross-sections which decrease as the available energy is increased. At very high energies these cancellations are absolutely necessary. Without them the sacred principle of conservation of probability would eventually be violated.

For the Fermilab experiments, a confirmation of the gauge theory structure thus corresponds to showing that the production rates for the vector boson pairs are the smallest possible for any choice of the vector boson vertices. Any rate enhancement at high energies would show the standard model to be faulty. Because of the probability conservation problems, this would imply that  $W$ s and  $Z$ s cease to be pointlike particles at this high energy scale  $E$  or, using Heisenberg's uncertainty relation, at the equivalent short distances of order  $\hbar c/E$ .

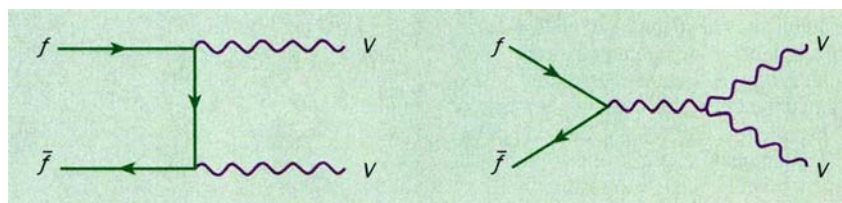
In their analysis of  $WW$  and  $WZ$  production<sup>3</sup>, the CDF Collaboration has searched for one vector boson decaying into two quarks (which would be seen as two jets of particles in the detector) and the other one decaying into electrons or muons. In an energy range around 500 GeV one such event was found where, in the absence of a  $WWZ$  vertex, 15 would have been expected. Thus, the existence

of non-vanishing interactions between  $W$ s and  $Z$ s has been established. Similarly, the D0 Collaboration has seen one event where a  $W^+W^-$  pair is produced and subsequently each of these  $W$ s decays into an electron (or positron, depending on sign) and a neutrino,  $\nu$ .

In the  $W$  plus photon channel both collaborations have seen about 25 events, at relatively low energies and thus probing correspondingly larger distance scales. Many of these events are probably due to  $W$  radiative decay, for example  $W \rightarrow e\nu\gamma$ , but at least ten clearly correspond to  $W\gamma$  production. These observations are, again, completely consistent with the gauge theory expectation. The determination of the  $W\gamma$  production cross-section corresponds to a measurement of the non-abelian couplings of three vector bosons with an error of about 30 per cent for at least some of the  $W$  multipole moments. Thus, we now have a direct confirmation that the intermediate vector bosons are indeed gauge bosons.

How do these measurements compare to the much more precise results which have been collected in electron-positron collisions at CERN over the past six years? At CERN well over ten million  $Z$ -decays have been analysed so far. They confirm the gauge theory predictions for the interactions of the  $Z$ -boson with quarks and leptons at the few parts in a thousand level. This stunning agreement has convinced nearly everybody that  $W$ s and  $Z$ s are indeed gauge bosons. Thus any report of a significant discrepancy in these

## Measuring the vertices



THE weak interactions between nuclei are caused by the exchange of the so-called weak bosons  $W$  ( $W^+$  and  $W^-$ ) and  $Z$ .  $W$ -exchange is responsible for the beta-decay of nuclei.  $W$ s and  $Z$ s carry spin 1 (in units of the reduced Planck constant  $\hbar = 1.05 \times 10^{-34}$  J s) and are therefore also called vector bosons. Within the standard model of particle physics they arise as gauge bosons of an underlying gauge symmetry group,  $SU(2) \times U(1)$ . The fourth partner in this quartet is the photon, the gauge boson of electromagnetism. The symmetry group  $SU(2)$  is non-commutative, or non-abelian, so the gauge bosons are said to have non-abelian character: they can directly couple to one another.

Two contributions determine the annihilation rate of a fermion and an anti-fermion into two vector bosons  $V$ , in other words the cross-section for the process  $f\bar{f} \rightarrow VV$ . The first graph describes radiation of the vector bosons off the incident particles and is solely responsible for processes like  $e^+e^- \rightarrow \gamma\gamma$ . In processes such as  $W^+Z$  production via annihilation between an 'up' ( $u$ ) and an 'anti-down' ( $\bar{d}$ ) quark, the second contribution corresponds to the decay of a virtual  $W^+$  into a real  $W^+$  and a  $Z$  boson. This contribution depends on the coupling of two  $W$ s and a  $Z$ . The process  $u\bar{d}$  to  $W^+Z$  therefore allows for a direct measurement of this  $WWZ$  'vertex'.

D. Z.

new, more direct measurements at Fermilab would have led to great consternation among theorists. It is reassuring to know that the three-vector-boson couplings follow the expected pattern as well.

Nevertheless, there remains the possibility that quantum corrections induced by hitherto unobserved, heavy particles lead to more subtle deviations from the gauge theory structure of the three-vector-boson vertices. Such deviations would be energy-dependent, becoming largest at energies of order  $E=Mc^2$  corresponding to the mass  $M$  of the heavy particles. Depending on the value of  $M$  and on the strength with which the new particles couple to  $W$ s,  $Z$ s and photons, one might or might not observe the large rate increases in production of vector boson pairs which have been searched for in the Fermilab experiments.

The strategy for measurements of the three-vector-boson vertices thus becomes twofold. The first strand is to extend the reach of the hadron collider experiments by collecting more data and becoming sensitive to sizeable enhancements of the cross-section for vector boson pair production at higher energies. The current value for the magnetic moment of the  $W$  (0.8 to 3.11 at the 95% probability limit, neglecting correlations) was obtained from data taken in 1992–93. Since then Fermilab's Tevatron has been running beautifully (leading, for example, to the discovery of the top quark) and the total amount of data has been increased by more than a factor of five. Once these data are analysed a substantially improved measurement of three-vector-boson vertices will result. But because of the limitations inherent to hadron colliders the sensitivity of these measurements will always depend on sizeable increases in the production cross-sections for vector boson pairs, of 20–50% or more, in particular at very high energies. More subtle effects would elude these experiments.

More precise measurements, albeit at smaller energies, are the domain of electron-positron colliding beam facilities. Here the Large Electron Positron collider (LEP) at CERN will come into the game in 1996, when its available energy is pushed above the  $W^+W^-$  pair production threshold. This second phase of LEP, known as LEP2 and following the first phase which was devoted to producing  $Z$  bosons, is expected to collect several thousand  $W^+W^-$  pairs in each of four experi-

ments. This will allow cross-section measurements with a precision approaching 1%, so reducing the error in the measurement of the three-vector-boson vertex by well over an order of magnitude. A deviation of a few per cent from the gauge theory value  $g_W=2$  would be observable in the LEP2 experiments.

The LEP2 measurements will be done at relatively low energies, where the large gauge theory cancellations which are characteristic for production of high-energy

vector boson pairs are not yet fully operative. Hence the measurements from LEP2 and the Tevatron will nicely complement each other. Combined, they will finally allow a much more precise answer to the question of how well the gauge theory predictions for the three-vector-boson couplings are indeed realized in nature. □

Dieter Zeppenfeld is in the Department of Physics, University of Wisconsin, Madison, Wisconsin 53706, USA.

## ION CHANNELS

# Annexins taken to task

Stephen E. Moss

ANNEXINS have long been proteins in search of secure employment, proposed functions for them having been presented even more frequently than new members of the family. The problem now is one of distinguishing fact from fiction within a choice of attractive possibilities. These include roles for annexins as cell-adhesion molecules, growth factors, receptors for

conventional view of how annexins interact with membranes.

The annexin family consists of at least ten genes in mammals, with further members in lower eukaryotes. All annexins have an internally repetitive structure, comprising four or eight repeats of a conserved 70-amino-acid domain, the most significant differences being in the length and composition of the amino-terminal domains. The presence of phosphorylation sites, alternatively spliced domains and ligand-binding sites in the amino-termini of several annexins suggests that these regions are probably responsible for defining distinct functions for members of the family. A defining feature of annexins is their ability to bind, courtesy of calcium ions, to negatively charged phospholipids such as those that enrich the inner leaflet of the plasma membrane and the cytoplasmic aspect of various organelles. It is therefore relatively easy to imagine annexins participating in the movement of vesicles about the cell, but how they might function as ion channels is less apparent.

Annexins V and VII can both form voltage-sensitive ion channels in artificial membranes<sup>3,4</sup> — the annexin V channel is permeable to several divalent cations whereas annexin VII is selective for  $Ca^{2+}$ . Support for the idea that annexin V might be a  $Ca^{2+}$  channel came from analysis of its crystal structure by Huber and co-workers, which revealed a doughnut-shaped molecule with a central pore lined by acidic amino acids<sup>5</sup>. On the other hand, a channel must, *ipso facto*, span the plasma membrane, and annexin V is found free in the cytosol (*a* in Fig. 2). Huber and his colleagues countered this objection with the sugges-

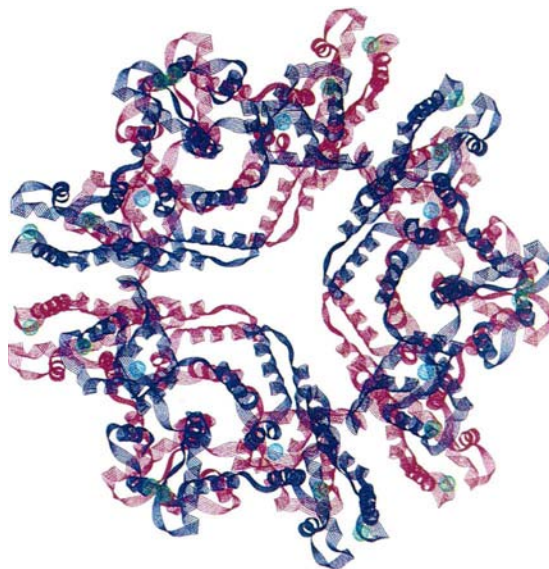


FIG. 1 Ribbon diagram of annexin XII (ref. 2). Two trimers (red and purple) lie face to face, forming a disc with a central pore. Basic residues lining the pore suggest that annexin XII functions as an anion channel.

viruses, protein kinase C inhibitors and receptors, anti-inflammatory agents, anti-coagulants, endocytic regulators, secretory regulators and calcium channels<sup>1</sup>.

Although it is unusual for proteins of unknown function to be scrutinized by structural biologists, the value of crystallography as a tool to determine function is borne out in a study described on page 512 of this issue<sup>2</sup> by Luecke *et al.* Their discovery of a new hexameric structure for annexin XII (Fig. 1) not only gives credence to annexins as ion channels and vesicle fusogens, but also challenges the

1. Abe, F. *et al.* (CDF Collaboration) *Phys. Rev. Lett.* **74**, 1936–1940 (1995).
2. Abachi, S. *et al.* (DO Collaboration) *Phys. Rev. Lett.* **75**, 1034–1039 (1995).
3. Abe, F. *et al.* (CDF Collaboration) *Phys. Rev. Lett.* **74**, 1941–1945 (1995).
4. Abachi, S. *et al.* (DO Collaboration) *Phys. Rev. Lett.* **75**, 1028–1033 (1995).
5. Abachi, S. *et al.* (DO Collaboration) *Phys. Rev. Lett.* **75**, 1023–1027 (1995).
6. Abe, F. *et al.* (CDF Collaboration) *Phys. Rev. Lett.* **75**, 1017–1022 (1995).



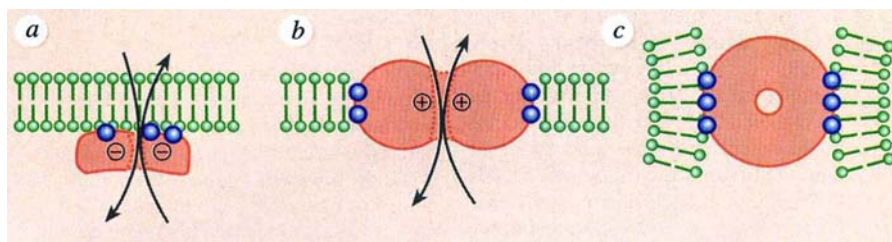


FIG. 2 Models of membrane interaction by annexins. *a*, Binding of monomeric annexin V to the phospholipid bilayer is mediated by calcium ions (blue spheres) at the slightly convex hydrophobic surface of the molecule. The acidic amino acids that line the central pore (dotted lines) are represented by minus signs. *b*, Hexameric annexin XII is fully integrated into the phospholipid bilayer. The interaction is stabilized by peripheral  $\text{Ca}^{2+}$  ions and the central pore is lined with positively charged amino acids. *c*, A possible way in which an annexin II hexamer (viewed from above or below) might simultaneously bind via  $\text{Ca}^{2+}$  to two separate phospholipid vesicles. Arrows indicate the direction of ion flow.

tion that annexin V might 'electroporate' the membrane during depolarization or hyperpolarization<sup>6</sup>, but failure to demonstrate that annexin V has  $\text{Ca}^{2+}$ -channel activity in intact cells has allowed the debate to rumble on.

The hexameric structure of annexin XII described by Luecke *et al.* provides the most plausible answer yet to the problem of how annexins span the plasma membrane (*b* in Fig. 2). In the disc-shaped hexamer, two annexin trimers join face to face, the interaction being stabilized by electrostatic bonding and the ligation of six shared calcium ions. In this model of membrane integration, a further 18 calcium ions lie in a ring around the periphery of the disc where they mediate binding to the phospholipid bilayer. Crucially, the annexin XII hexamer not only has the requisite spatial dimensions of a membrane-spanning protein, it also has a central pore lined with charged residues. A model of annexin V based on the annexin XII hexamer would also have a central pore lined with charged residues, but with the key difference that in annexin XII the residues are basic whereas in annexin V they are acidic. The obvious implication is that annexins are able to form both anion- and cation-specific channels.

A remarkable feature of the annexin hexamer is that it not only supports the idea that annexins are ion channels, it also provides an explanation for the ability of some annexins to act as vesicle fusogens. Annexin II is implicated in both exocytosis<sup>7</sup> and endocytosis<sup>8</sup>, but how can a monomeric annexin with its single hydrophobic face, buried with the help of

$\text{Ca}^{2+}$  ions into one membrane, bring two opposing vesicle membranes together? Annexin II may overcome this problem in two ways, either as part of a heterotetramer or as a homohexamer. In most cells there is a pool of annexin II heterotetramers, in which two molecules of annexin II are linked through two molecules of a small protein, p11. It is (conceptually) not surprising that an annexin II heterotetramer can stimulate secretory granule aggregation, but it is less easy to see how monomeric annexin II might mediate vesicle aggregation and fusion. A hexameric form of annexin II ringed by sticky calcium ions, similar to that described by Luecke *et al.* for annexin XII, now offers a satisfactory structural

#### TAXONOMY

## The species alias problem

Robert M. May and Sean Nee

How many living species have been named and recorded? We do not know. For most groups, the information exists in scattered databases, much of it still on file cards. Estimates of the total run at around 1.7–1.8 million<sup>1,2</sup>. But this count is inflated by synonyms: single species that have been separately named and recorded on two or more occasions. Given that, for instance, some 40% of all 300,000 or so named beetle species are known from only one geographical site, and that no synoptic database exists, this synonymy problem should not surprise us.

Careful studies of particular groups, mainly insects, have typically found 'synonymy rates' of around 20%; that is, one-fifth of the names are 'aliases', although for some groups the rate can exceed 50% (ref. 3). On this basis, E. O. Wilson and others have corrected downwards, suggesting a global total of around 1.4 million named living species<sup>4</sup>. The rate, however, tends to vary quite a bit from group to group. More importantly, any such rate derived from known synonyms must be a

explanation for fusogenic activity (*c* in Fig. 2).

The discovery of the annexin hexamer undoubtedly adds weight to the idea that annexins function as ion channels and vesicle fusogens, but the new structure provokes several questions. How does hexamer formation take place? Clearly, there is a requirement for  $\text{Ca}^{2+}$  at the interface between the two opposing trimers and, given the known affinity of annexins for  $\text{Ca}^{2+}$ , resting cytosolic  $\text{Ca}^{2+}$  is likely to be too low to allow multimerization to occur. How does the hexamer insert into the membrane? Would the complete structure form within the cytosol and then somehow integrate into the lipid bilayer, or could opposing trimers meet from opposite sides of the membrane? And most crucially, how would an annexin ion channel be gated?

*In vitro*, annexin V functions as a  $\text{Ca}^{2+}$  channel in response to both hyperpolarizing and depolarizing potentials, but there is no evidence to show that it forms a voltage-gated  $\text{Ca}^{2+}$  channel in whole cells. The challenge now is to demonstrate that annexins function as ion channels or fusogens in intact cells. The design of experimental strategies stimulated by the existence of an annexin hexamer should lead to some interesting answers. □

Stephen E. Moss is in the Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK.

1. Raynal, P. & Pollard, H. B. *Biochim. biophys. Acta* **1197**, 63–93 (1994).
2. Luecke, H., Chang, B. T., Maillardi, W. S., Schlaepfer, D. D. & Haigler, H. T. *Nature* **378**, 512–515 (1995).
3. Rojas, E., Pollard, H. B., Haigler, H. T., Parra, C. & Burns, A. L. *J. Biol. Chem.* **265**, 21207–21215 (1990).
4. Burns, A. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 3798–3802 (1989).
5. Huber, R., Römisch, J. & Paques, E. *EMBO J.* **9**, 3867–3974 (1990).
6. Karshikov, A. *et al. Eur. J. Biophys.* **20**, 337–344 (1992).
7. Creutz, C. E. *Science* **258**, 924–931 (1992).
8. Gruenberg, J. & Emans, N. *Trends Cell Biol.* **3**, 224–227 (1993).

lower limit, with other synonyms yet to be uncovered.

Solow, Mound and Gaston have now made a start on estimating the true rate of synonymy<sup>5</sup>. They use the records of thrip (Thysanoptera) species, as published each year since 1901. Altogether, there have been 197 workers who have named thrip species and 28 of these have had all their names synonymized. Mound has personal experience of the synonymy problem, having discovered that one of the many insect species he has named had previously been given another name — by himself!

Out of the total of 6,112 thrip names, 1,326 are currently known to be synonyms, giving an observed synonymy rate of 22%. But we also know what proportion of the names published in each year are now known to be synonyms. As might be expected, there is a higher rate for the names given in earlier years, reflecting the fact that it takes time to uncover aliases. This provides the opportunity to fit a probability distribution to the time taken to uncover a synonym, and so to estimate

# IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Males and females of the Eclectus parrot, *Eclectus roratus*, have such different coloration that for many years they were thought to be separate species (the male is green). The fact that the two sexes have the same coloration in most parrot species makes this an even more striking example of the widespread problem of synonymy — the existence of two or more names for the same species.

how many have yet to be revealed. Using the gamma distribution (a flexible two-parameter distribution) as their model of the amount of time it takes to reveal a synonym, Solow *et al.* employ likelihood techniques to estimate that the true proportion of synonyms is around 39%, about double the observed rate. They also estimate that, on average, it takes 43 years to identify a synonym.

As a problem in statistical estimation, this particular example belongs to a large class which includes such diverse questions as estimating the size of an author's vocabulary from observed samples of writing, or the true number of executions in South Vietnam from official records<sup>6</sup>. When placed in such company, the analysis of Solow *et al.* stands out as particularly simple in its detail. Their method may, in fact, be the simplest possible likelihood model for problems of this kind. We thought that an even simpler approach might be to use the one-parameter exponential distribution, instead of the two-parameter gamma, to model the time to detection. But the constraints on the database (in particular, the curve of observed synonymies against time is necessarily 'anchored' to zero in the last time period of the study) are such that this approach turns out to be excessively simple, producing a nonsensical answer: with this one-parameter distribution, we are led to the conclusion that all names are synonyms, and that it will take more than 200 years to discover the fact. The exercise, however, serves to underline the technical difficulties that can arise when trying to estimate true synonymy rates.

Hans Reinhard/Bruce Coleman Ltd

In fact, such questions of how many words Shakespeare knew (akin to the true synonymy rate, and thus the true total number of recorded species) turn out to be more difficult to answer than questions of how many new words might be expected in a newly discovered sonnet by Shakespeare<sup>7</sup> (akin to the observed synonymy rate). The latter is, in principle, a testable prediction. We hope it will be possible to test the implicit predictions of Solow and colleagues' model by noting how many synonyms will be discovered in, say, the next two years. The outcome of such tests will tell us more about the many issues that naturally arise from their pioneering paper.

Apart from the statistically technical questions about drawing conclusions from the available data for one group, such as thrips, there are other questions about how representative thrip synonyms may be. Rates of synonymy are lower among the more charismatic groups, such as birds and mammals, which have always received more attention than most invertebrates (but see illustration). And among invertebrates, we can be fairly sure that butterflies, and even thrips, have been more favoured by taxonomists than, say, nematodes. Suppose, however, that Solow and colleagues' upward revision from an observed synonymy rate among thrip names of around 20%, to a true synonymy rate of around 40%, does indeed prove to be generally representative. The implication would be that the true global total of named and recorded species could be around one million.

The total number of species alive today is, of course, even more uncertain, with estimates ranging from 3 million to 30 million or more. Many, though by no means all, of these estimates are based on the number of distinct species currently named and recorded. So any downward revision in this number, as a result of better understanding of synonymy rates, will at the same time tend to lower estimates of the global species total. □

Robert M. May and Sean Nee are in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

1. May, R. M. *Phil. Trans. R. Soc. B* **330**, 292–304 (1990).
2. Stork, N. E. *Biol. J. Linn. Soc.* **35**, 231–237 (1988).
3. Gaston, K. J. & Mound, L. H. *Proc. R. Soc. B* **251**, 139–142 (1993).
4. Wilson, E. O. *The Diversity of Life* (Penguin, London, 1992).
5. Solow, A. R., Mound, L. A. & Gaston, K. J. *Syst. Biol.* **44**, 93–96 (1995).
6. Bunge, J. & Fitzpatrick, M. J. *Am. Stat. Ass.* **88**, 364–373 (1993).
7. Efron, B. & Tibshirani, R. *Biometrika* **63**, 435–447 (1976).

DAEDALUS

## Recolonization

DELHI belly, Montezuma's revenge, gippy tummy and so on, are all jocular terms for traveller's diarrhoea. It is often blamed on poor foreign hygiene, but this is unfair; travellers to the most sanitary countries can get it. The traveller's gut organisms, carried from his home region, find themselves challenged by the local varieties. These are better adapted to local conditions (which is why they are established there in the first place). They displace the traveller's own organisms, stirring up in the process a volcanic intestinal immune response. When it dies down, the traveller is as immune to the locals as he was to his previous inhabitants.

Daedalus is now at work on the obvious countermeasure — immunization. DREADCO biologists are collecting samples of indigenous gut organisms from all over the world, and are drawing up a global map of their distribution. The samples are being dried, irradiated and so on, to produce enfeebled and attenuated forms which cannot multiply, but still put the immune system on guard against them. When the system has been perfected, a traveller will simply tell DREADCO where he plans to go, and will receive a capsule of attenuated gut organisms from that region. He will take it about ten days before his flight. As with many immunizations, he may experience brief mild discomfort, but nothing dramatic. When he reaches his destination, the local organisms will recolonize his colon quietly and without trouble. He will be immune to them already.

If he stays a month or longer, he may suffer when he gets back; his immune system will have forgotten its defences against his home organisms. He will be well advised to take a DREADCO 'safe return pill' ten days before his homeward journey. It will immunize him against his old indigenous inhabitants.

DREADCO's growing world map of intestinal organisms has an interest of its own. Why do particular strains dominate certain regions? They must be best adapted to the local conditions, which from their point of view means the local food. Daedalus expects their distribution to conform closely to the map of culinary traditions. Rice, olive oil, yams, wheat, all must encourage their own gut organisms. Those insular British travellers who take a fortnight's supply of sandwiches with them are shrewdly preventing foreign bugs from penetrating their defences. Conversely, when in due course hamburgers and Coca-Cola have conquered the world, and the same diet prevails over the entire planet, traveller's diarrhoea will fade away. David Jones

# Underwater quantum coding

**SIR**—Quantum mechanics is often presented as a limitation on achieving certain tasks, such as the simultaneous determination of position and momentum. Recently, the idea of using Heisenberg's uncertainty relations for secure communications has been explored, both theoretically (see, for example, ref. 1) and experimentally (for example, refs 2, 3). Here we report a cryptographic channel using a 23-km optical cable below Lake Geneva.

The idea is that whenever a third party taps into a quantum communication channel, this action necessarily perturbs the signal, provided that the bits of the key are represented by incompatible quantum states. This unavoidable perturbation can then be detected by controlling the correlation between the transmitted and the

received signals. No information is transmitted; only a coding key is established. Once the secrecy of the key has been checked, it can safely be used to encode the message using standard cryptography<sup>4</sup>. The future of quantum cryptography depends on the possibility of using standard telecommunications fibres over some tens of kilometres<sup>5,6</sup>, and on the evolution of current public key cryptosystems whose safety is still not proven mathematically<sup>7</sup>.

In our experiment, the quantum channel was one of 40 fibres installed in a cable operated by Swiss PTT at 2.5 Gbit s<sup>-1</sup> between Nyon and Geneva. Most of the cable is under water; its attenuation is 8.2 dB and its length 22.7 km. We also performed tests in the laboratory with a 26-km-long fibre coil. A 1,300-nm pulsed laser with 1 ns pulse width and 1.1 MHz pulse rate was used. The key was encoded using the polarization of light pulses manually selected by a rotating polarizer placed after a  $\lambda/4$  wave plate. A polarization controller, compensating the polarization modification due to the fibre link, was used to align the emission and measurement axis.

The bit error rate (BER), the ratio of the number of wrong detections to the total number of detections, is the sum of the detector noise and the optical noise. We obtained a BER of 4.3% in the laboratory and 3.4% in the field, with an error of 0.3% (see figure), using low-intensity pulses. The optical noise consists of depolarization due to polarization-mode dispersion and polarization fluctuation due to birefringence and topological effects (see refs 2, 8–10 for details). A polarization separation of 23 dB has been obtained using an intense continuous-wave light source, both in the laboratory and in the field. This corresponds to a  $BER_{opt}$  of 0.5%. The stability of the polarization alignment was excellent for most of the time. When not stable, a fast polarization controller could overcome larger fluctuations.

The detector noise ( $BER_{det}$ ) can be evaluated as follows:

$$BER_{det} = (D\delta\tau)/(D\delta\tau + (p_{inj}L\eta))$$

where  $\delta\tau$  is the detection time window,  $D$  is the number of counts per unit time not due to incoming photons (700 c s<sup>-1</sup>),  $p_{inj}$  is the number of photons per pulse at the transmitter output (0.12 per pulse),  $L$  is the sum of the losses in the transmission cable and in the receiver's polarization splitter ( $\sim 10$  dB), and  $\eta$  is the overall quantum efficiency of the detector (0.2%). This corresponds to a  $BER_{det}$  of around 3%. Thus, the BER is mostly caused by the detector and is low enough for standard algorithm of errors correction and privacy

amplification to be applied, and to guarantee the privacy of the key.

We have thus demonstrated the feasibility of a quantum cryptographic system using polarization coding in the second telecommunications window around 1,300 nm on installed optical cables. The measured bit error rate (3.4%, before any error correction) is low enough to guarantee the privacy of the quantum communication channel. Consequently, the issue is no longer whether the quantum technology works, but whether it can be made robust and practical enough for use over an optical network.

**A. Muller, H. Zbinden**

**N. Gisin**

Group of Applied Physics,  
University of Geneva,  
1211 Geneva 4, Switzerland

1. Ekert, A. *Nature* **358**, 14–15 (1992).
2. Muller, A., Breguet, J. & Gisin, N. *Europhys. Lett.* **23**, 383–388 (1993).
3. Franson, J. D. & Jacobs, B. C. *Electron. Lett.* **31**, 232–234 (1995).
4. Shannon, C. E. *Bell Syst. tech. J.* **28**, 656 (1949).
5. Hughes, R. J., Luther, G. G., Morgan, G. L. & Simmons, C. *Proc. 7th Rochester Conf. Coherence and Quantum Optics* (1995).
6. Marand, C. & Townsend, P. D. *Opt. Lett.* **20**, 1695–1697 (1995).
7. Grollman, J. & Selman, A. L. *SIAM J. Computing* **17**, 309–335 (1988).
8. Gisin, N. & Pelloux, J. P. *Opt. Commun.* **89**, 316–323 (1992).
9. Tomita, A. & Chiao, R. Y. *Phys. Rev. Lett.* **57**, 937–940 (1986).
10. De Angelis, C., Galtarossa, A., Gianello, G., Matera, F. & Schiano, M. J. *Lightwave Technol.* **10**, 552–555 (1992).

## Ambient ozone and loblolly pines

**SIR**—McLaughlin and Downing<sup>1</sup>, studying loblolly pine (*Pinus taeda* L.) trees, conclude that “ozone exposures at  $>40$  nl l<sup>-1</sup> interacted with low soil moisture and high air temperatures to reduce short-term rates of stem expansion. Annual growth was also inversely related to seasonal ozone exposure and soil moisture stress.” We suspect that most of the within-year (short-term) variation in stem growth is accounted for by the growth trend term (slope against time) in their model, leaving relatively little to be explained by ozone or other environmen-

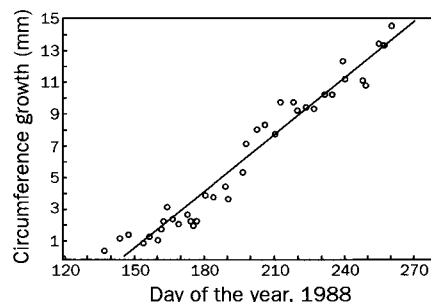
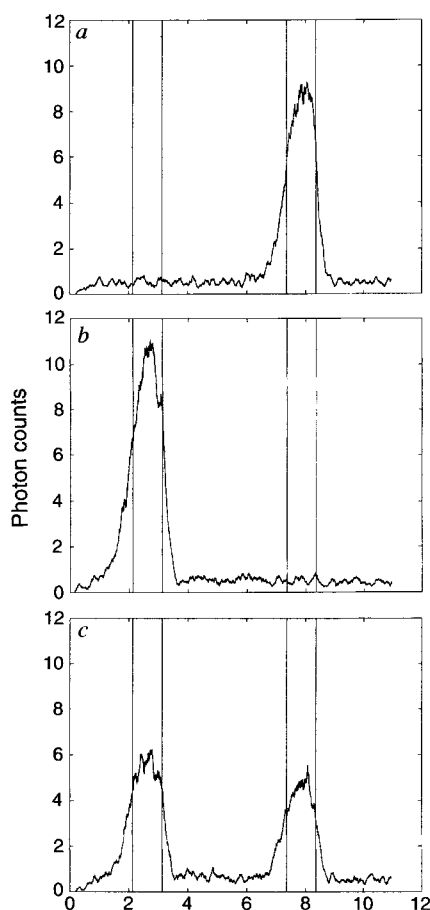


FIG. 1 Plot of simple linear regression against observed values (circles) for tree 23. Transcribed from McLaughlin and Downing's Fig. 2.



Time histogram of the detected photons in three different configurations. The horizontal scale gives the time in nanoseconds and the vertical scale the corresponding number of detected photons. Photons are 90° polarized (a), 0° polarized (b) and 45° polarized (c), in order to simulate the selection of a wrong detection basis or the effect of an eavesdropper. The number of wrong counts is very small in a and b, in contrast to c, where the incoming photons are found to be distributed equally.

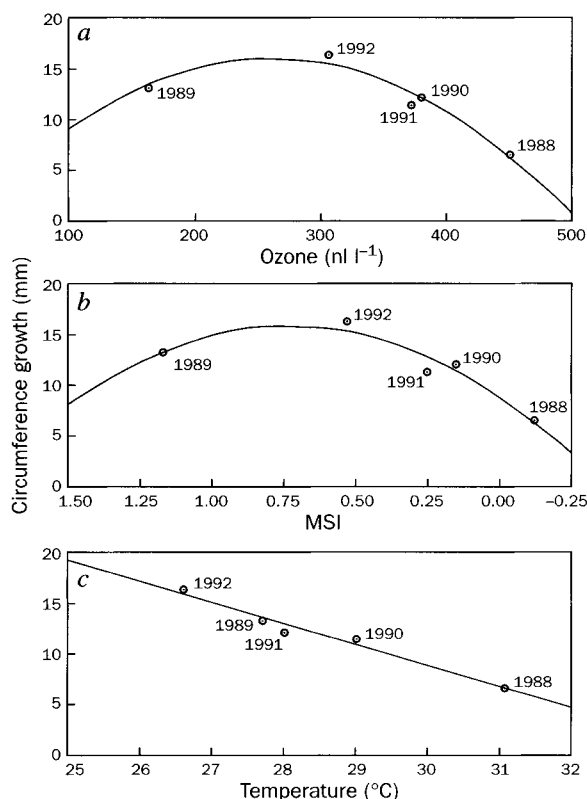


FIG. 2 a, Annual growth against ozone from McLaughlin and Downing's Fig. 3b. b, Annual growth against moisture stress (MSI). c, Annual growth against temperature. For b and c, annual growth data are transcribed from McLaughlin and Downing's Fig. 3b; MSI and temperature are from their Table 1.

tal variables. To test this idea, we transcribed the data from McLaughlin and Downing's Fig. 2 for tree 23 and fitted a straight line (stem growth =  $b_0 + b_1(\text{day}) + e_t$ , where  $b_0$  and  $b_1$  are regression coefficients and  $e_t$  is a random error term). The estimated coefficients (equivalent to slope against time) account for 95% of the variation, leaving only 5% to be explained by other variables (Fig. 1).

Within-year variations in stem growth not explained by the annual growth trend term can be modelled as a response to ozone acting in combination with temperature and/or moisture stress. We note, however, that the relationship between annually detrended growth and ozone is inconsistent, and that it is significant for >50% of the trees in only 3 of 5 years, and not significant for any of the trees in 1989<sup>1</sup>. The statistical significance of ozone interactions should be interpreted with caution, given the nonsignificance of ozone as a main effect<sup>2</sup> and the fact that high levels of ozone and hot dry weather often occur together<sup>3,4</sup>.

McLaughlin and Downing suggest that stem growth variation is attributable to the synergistic action of ozone and water stress over periods of a few weeks. This conclusion is unexpected, because many studies have found that short-term effects of ozone are greater in well-watered plants than in plants subjected to water

stress<sup>5</sup>. Apparent negative growth rates in many periods shown in their Fig. 1 are attributable to normal hydraulic stem shrinkage during periods of high transpiration<sup>6</sup>. This suggests that their high-frequency growth model is based on short-term variations in stem moisture content. McLaughlin and Downing also modelled between-year variation in total annual stem growth as a response to ozone. In their Fig. 3, they show a curvilinear relationship in which annual stem growth decreases at the highest levels of seasonal ozone exposure. Unambiguous interpretation of this relationship is problematic, because the effect of ozone on total stem growth is confounded with the effects of moisture stress and temperature. Illustrative of this is a plot of average annual stem growth against ozone, moisture stress and temperature (Fig. 2). The ozone and moisture stress curves are almost identical.

We share McLaughlin

and Downing's interest in forest responses to tropospheric ozone, and commend their effort to develop a new research approach. Unfortunately, their approach has serious limitations which should be addressed before they are applied further. We do not believe the data and analyses in McLaughlin and Downing's paper support their inference about the effects of ambient ozone on the loblolly pine.

**Gregory A. Reams**

*Institute for Quantitative Studies,  
USDA Forest Service,  
Southern Research Station,  
New Orleans, Louisiana 70113, USA*

**Paul C. Van Deusen**

*National Council for Air and Stream  
Improvement,  
Tufts University, Medford,  
Massachusetts 02155, USA*

**Alan A. Lucier**

*National Council for Air and Stream  
Improvement,  
PO Box 13318, Research Triangle Park,  
North Carolina 27709, USA*

**McLAUGHLIN AND DOWNING REPLY** — Reams *et al.*'s comments misrepresent several aspects of our analyses of loblolly pine growth patterns<sup>1</sup>. First, they attempt to generalize relative influences of time and environment on overall growth rates using a single tree (tree 23, see Fig. 1), one we identified as a fast-growing indi-

vidual resistant to the episodic growth suppressions noted in most other trees. Although linear growth trends explained 97.5% of the variability in tree 23 over time, they explained less than 80% of the variability for all trees and less than 45% and 1% of the variation observed in trees 3 and 5, respectively. Time trends for all trees were first removed in our analyses, allowing high-frequency variance to be analysed separately, a well-established approach in dendroclimatic analyses<sup>7</sup>.

Reams *et al.* also confuse variations in influences of ozone on growth, across trees and time, with inconsistency. Variability in ozone effects among individual loblolly pine families and trees within families is well established<sup>8,9</sup>. Dose-proportional variations in observed responses over time, including a lack of detectable effects during a very-low-ozone year (1989), should be considered as evidence of consistency, not inconsistency.

Contrary to their assertion, direct effects of ozone were measured on short-term growth during the years of highest ozone exposure (see Fig. 3). Significant interactions among ozone and moisture stress represented in the multi-year model were more typical, and were amplified at higher ozone and moisture stress levels. One should not infer that either ozone or moisture stress effects were less valid because they were represented only by interaction terms in the multi-year model.

Reams *et al.* are surprised at the enhancement of ozone effects by water stress in mature trees. The studies they cite on 120-day-old aspen seedlings<sup>5</sup> concluded that drought apparently caused stomatal closure, temporarily protecting these small seedlings from ozone uptake and damage. Those authors also reported that post-drought effects of ozone on biomass were greater than those of drought alone, and hypothesized that drought would cause compensatory enhancement of ozone effects on mature trees.

There is good evidence that ozone enhances moisture stress<sup>10</sup>, including

1. McLaughlin, S. B. & Downing, D. J. *Nature* **374**, 252–254 (1995).
2. Neter, J. & Wasserman, W. *Applied Linear Statistical Models* (Irwin, Homewood, IL, 1974).
3. US Environmental Protection Agency EPA Report EPA-600/8-84-020aF-eF, 5v. (NTIS, Springfield, VA, 1986).
4. US Environmental Protection Agency EPA Report EPA/600/8-88/105F (NTIS, Springfield, VA, 1992).
5. Pell, E. J., Temple, P. J., Friend, A. L., Mooney, H. A. & Winner, W. E. *J. env. Qual.* **23**, 429–436 (1994).
6. Lassole, J. P. *Forest Sci.* **25**, 132–144 (1979).
7. Grissino-Mayer, H. D. & Butler, D. R. *Southeast. Geograph.* **1**, 65–81 (1994).
8. McLaughlin, S. B. *et al. J. env. Qual.* **23**, 247–257 (1994).
9. Taylor, G. E. *J. env. Qual.* **23**, 63–82 (1994).
10. Chappelka, A. H. & Freer Smith, P. H. *Envir. Pollut.* **105**, 105–117 (1995).
11. Maier-Maerker, U. & Koch, W. *Trees* **7**, 12–25 (1992).
12. McLaughlin, S. B. & Downing, D. J. *Can. J. For. Res.* (in the press).
13. Fox, S. & Mickler, R. *The Impact of Air Pollutants on Southern Pine Forests* (Ecol. Stud. Ser. 118, Springer, Berlin, 1995).



studies with mature trees in which ambient ozone increased transpirational loss of water<sup>11</sup>. In large trees, larger hydraulic reserves of water relative to leaf area allow ambient ozone to enhance moisture stress without causing stomatal closure.

Hydraulic shrinkage is normal in forest trees and is also routinely accompanied by slowing of stem growth<sup>6</sup>. Lasso showed that net basal area increment and stem shrinkage of Douglas fir were well related to daily moisture deficit<sup>6</sup>. Our analyses showing that ozone can significantly reduce stem increment or enhance shrinkage under drier conditions are consistent with induction and/or amplification of moisture stress.

Finally, Reams *et al.* point to the importance of annual-scale interactions among ozone, soil moisture stress and temperature. These all directly influence whole-tree moisture stress and growth; however, annual-scale comparisons of linear growth

trend coefficients over 5 years are not a robust analysis of those effects. For this reason, we did not attempt to identify an ozone-specific component of those low-frequency growth trends in our original analyses<sup>1</sup>. Important differences between wet and dry sites, and between short- and longer-term components of growth are not represented in Fig. 1, which plots averaged trend slopes, not total growth. In addition, apparent associations among these variables expressed at yearly scales can misrepresent physiologically significant relationships that occur at shorter time scales<sup>12</sup>. For example, temperature was very poorly related to both ozone exposure and growth in 1988 (Fig. 3). Our empirical model showed temperature effects to be negative only when combined with high ozone levels. Because of potentially misleading interactions at longer

time scales, short-term measurements and analyses can help to identify causative processes that control longer-term growth patterns.

There is now solid evidence from numerous studies<sup>13</sup> to substantiate the direct effects of ozone on growth of loblolly pine. We believe our analyses make an important beginning to understanding major aspects of those effects on mature trees. They offer significant possibilities for the forest industry to understand better a much wider array of environmental interactions that influence forest growth processes and forest production potential.

**S. B. McLaughlin**

**D. J. Downing**

*Environmental Sciences Division,  
Oak Ridge National Laboratory,  
Oak Ridge, Tennessee 37831, USA*

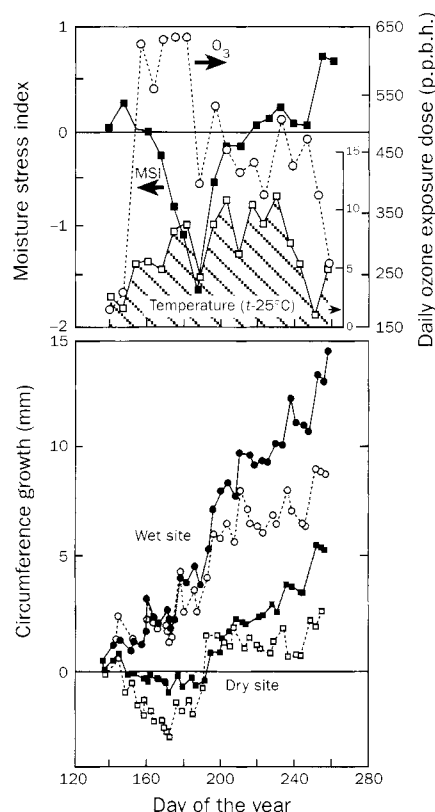
## Sisters like it hot

**SIR**—The sex of many reptile species is determined by the incubation temperatures that the animal experiences as an embryo<sup>1-4</sup>. Although this system has evolved from genotypic sex determination (GSD) several times, its adaptive significance remains controversial<sup>1-5</sup>. Here we show that incubation temperatures affect body size and performance differentially in male and female lizards, thereby offering a plausible adaptive advantage to temperature-dependent sex determination (TSD) over GSD. By linking sex determination to incubation temperatures, the size and performance of offspring may be maximized over a wide range of incubation conditions.

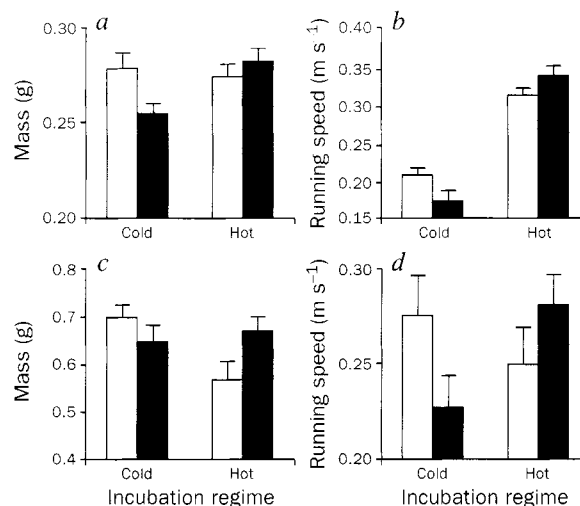
Charnov and Bull<sup>1</sup> suggested that TSD may evolve because it enables an embryo to develop as the sex best suited to its incubation environment. We tested the central prediction of this hypothesis — that there are sex-specific thermal optima for embryonic development — using a GSD lizard species (the Australian skink, *Bassiana duperreyi*), so that we could compare male and female offspring at each incubation temperature. We divided the eggs from each clutch between two incubators, so that we could compare siblings developing under incubation treatments that simulated the extremes of thermal regimes that we have measured in natural nests of this species (ref. 6, and R.S. and P.S.H., unpublished data) (means of 20 as

against 27 °C, with an 8 °C diurnal range). Egg sizes did not differ between the sexes, and sex ratios were unaffected by incubation temperatures, but sizes at hatching (both mass and snout-vent length) showed strong effects. Sons were larger from low-temperature incubation, whereas daughters were larger from high-temperature incubation (two-factor analysis of variance with incubation treatment and sex as the factors, and hatchling size as the dependent variable: mass interaction,  $F_{1,82} = 5.42$ ,  $P < 0.023$ ; snout-vent length interaction,  $F_{1,82} = 9.17$ ,  $P < 0.004$ ; see figure).

Performance (running speed at 20 °C



**FIG. 3** Variations in ozone, temperature, and moisture stress on a weekly time scale show distinctly different patterns across time. Growth of 27 of 28 trees was significantly reduced by ozone exposure during the 1988 season. Direct ozone effects were noted on 14 trees, while ozone interacted with moisture stress or temperature to reduce growth of 13 additional trees. Significant differences in growth responses occurred between faster- and more slowly growing trees and reflected differing sensitivities to environmental variables, of which ozone was an important component. Modified from McLaughlin and Downing<sup>1</sup> to include temperature data. ■, Tree 3; □, tree 5; ●, tree 23; ○, tree 28; p.p.b.h., parts per billion hours.



Experimental incubation at different thermal regimes differentially affects male and female skinks (*Bassiana duperreyi*). **a**, Body mass of hatchlings; **b**, running speeds at age 2 weeks; **c**, body mass at age 14 weeks; and **d**, running speeds at age 14 weeks. □, Male; ■, female. Sons from eggs incubated at colder temperatures are larger (**a**, **c**) and run faster (**b**, **d**) than their sisters. In contrast, higher-temperature incubation results in daughters that are larger and run faster than their brothers. These differences are significant from hatching, and persist for at least 14 weeks (**c**, **d**). Histograms show mean  $\pm$  1 s.e.

on a lizard raceway<sup>6</sup>) was also influenced by the interaction between incubation temperature and sex ( $F_{1,72} = 7.42$ ,  $P < 0.009$ ). Sons that hatched from cool-incubated eggs ran faster than daughters; but if the eggs had developed at higher temperatures, daughters ran faster than sons (see figure). These patterns were the same at 14 weeks of age. Although the young lizards more than doubled in mass over this period, the interaction effect between sex and incubation treatment remained significant both for body size (mass,  $F_{1,39} = 5.35$ ,  $P < 0.027$ ; snout-vent length,  $F_{1,39} = 10.47$ ,  $P < 0.003$ ) and running speed ( $F_{1,37} = 4.71$ ,  $P < 0.037$ ). These were independent effects, because incubation-induced differences in running speed remained significant ( $P < 0.05$ ) when size effects were removed from the analysis using analysis of covariance.

Because body sizes and running speeds are likely to affect the survival and eventual reproductive success of the animal<sup>7-9</sup>, our results suggest that the optimal incubation

temperature differs between males and females. This effect provides a plausible selective advantage for the evolution of temperature-dependent sex determination in reptiles: TSD may enable the embryo to develop as the sex best suited to the incubation regime it encounters<sup>1</sup>. It is, however, puzzling to see such strong effects in a GSD species, when theoretical models predict it only in TSD forms.

**Richard Shine, Melanie J. Elphick  
Peter S. Harlow**

*Biological Sciences A08,  
The University of Sydney,  
New South Wales 2006, Australia*

1. Charnov, E. L. & Bull, J. J. *Nature* **266**, 828–830 (1977).
2. Bull, J. J. *Q. Rev. Biol.* **55**, 4–21 (1980).
3. Janzen, F. J. & Paukstis, G. L. *Evolution* **45**, 435–440 (1991).
4. Rhen, T. & Lang, J. W. *Am. Nat.* (in the press).
5. Joanen, T., McNease, L. & Ferguson, M. W. J. in *Wildlife Management: Crocodiles and Alligators* (eds Webb, G. J. W. et al.) 533–537 (Surrey Beatty, Sydney, 1995).
6. Shine, R. *Am. Nat.* **145**, 809–823 (1995).
7. Fox, S. F. *Evolution* **29**, 95–107 (1975).
8. Ferguson, G. W. & Fox, S. F. *Evolution* **38**, 342–349 (1984).
9. Laurie, W. A. & Brown, D. J. *J. Anim. Ecol.* **59**, 529–544 (1990).

## Solution for the Sherborne problem

**SIR** — Since its discovery in quarry debris near Sherborne, Dorset<sup>1</sup>, the “Sherborne bone” has been the subject of debate, much of it in this journal (for example, refs 2, 3). We have re-studied this artefact, using optical microscopic analyses with image processing and a chemical and mineral textural study, followed by sampling for radiocarbon accelerator dating, to attempt to settle its authenticity once and for all.

This bone, a fragment of mammalian rib, had been engraved with the head and forequarters of a horse, and its resemblance to a Palaeolithic depiction from Creswell Crags, illustrated by Boyd Dawkins<sup>4</sup>, was used as evidence both for and against its authenticity. More recent exchanges (for example, refs 5, 6) have supported or opposed the authenticity

of the engraving as Palaeolithic, while Oakley, in Farrar<sup>5</sup>, reported relative dating analyses which indicated that the bone itself was fossilized.

Our analysis of the obverse of the fragment revealed that the spongy bone is still filled with sediment. Micro-roots present in the sediment are trapped in the trabeculae of the spongy bone. The sediment and roots are the residue of the original filling adhering to the bone when it was buried and are not the result of fraudulent additions to age a fresh bone artificially. Microanalysis by energy-dispersive X-ray spectrometry, and elemental mapping of the sediment filling the spongy bone and of that still adhering to the engraved side, showed them to be of similar composition, suggesting that the engraved side was also not artificially patinated. The patina covering the bone is therefore the result of the burial environment of the bone fragment.

Analysis of the engraving, however, revealed that almost all the engraved lines are sediment-free and do not show the same patina as the remaining surface of the bone. This is confirmed by optical analysis indicating that engraved lines have gray-value histograms that are different from those obtained from unengraved areas, but similar to those of recently damaged surfaces. Sediment residue also covers

eroded areas, suggesting that alteration of the bone surface took place before its engraving.

The engraved lines reveal none of the features that are generally visible on experimental lines produced by lithic tools on fresh bone<sup>7</sup>, such as sharp edges and multiple parallel striae. In contrast, the engraved surface of the Sherborne bone displays a granular, rough texture, and fractures perpendicular to the groove direction. The edges of the main grooves are frayed by continuous microflaking of the surface lamellae, clearly showing that the engraving took place on an already weathered bone.

Samples for radiocarbon dating were taken from the uncleaned and non-engraved obverse of the rib. The surface was mechanically cleaned and a small quantity (250 mg) of bone removed by drilling. After chemical pretreatment<sup>8</sup> and combustion<sup>9</sup>, the sample yielded an accelerator age (OxA 5239) of  $610 \pm 45$  years, indicating (after calibration<sup>10</sup>) that the rib had come from an animal that had died some time between the end of the 13th and the start of the 15th centuries AD. It is not possible to say when, after this date, the engraving was carried out, but it now seems inescapable that the Sherborne engraving is a recent fake. It is even possible that the horse head was traced by a metal tool, as no proof of the use of a flint point, such as the presence of minute striations accompanying the main groove<sup>7</sup>, was found. Oakley's determination that the rib was “fossilized”<sup>5</sup> can be attributed to the known limitations of relative dating techniques<sup>11</sup>.

**C. B. Stringer**

*Department of Palaeontology,  
The Natural History Museum,  
London SW7 5BD, UK*

**F. d'Errico**

*Institut du Quaternaire,  
33405 Talence, France*

**C. T. Williams**

*Department of Mineralogy,  
The Natural History Museum,  
London SW7 5BD, UK*

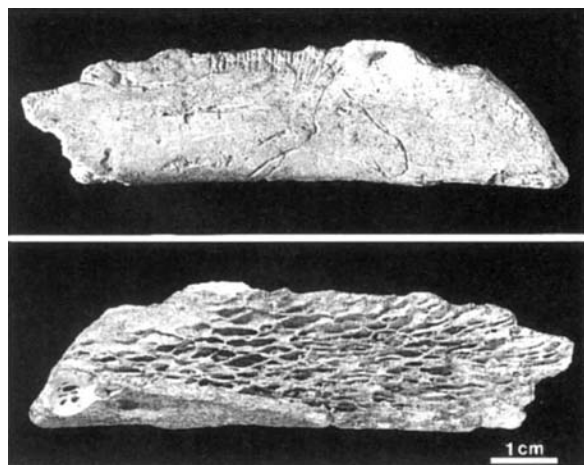
**R. Housley**

*Department of Archaeology,  
University of Glasgow,  
Glasgow G12 8QQ, UK*

**R. Hedges**

*Radiocarbon Accelerator Unit,  
Oxford OX1 3QJ, UK*

1. Woodward, A. S. *Q. J. geol. Soc.* **70**, 100–102 (1914).
2. Woodward, A. S. *Nature* **142**, 86 (1926).
3. Sollas, W. J. *Nature* **142**, 233 (1926).
4. Dawkins, W. B. *Early Man in Britain* (Macmillan, London, 1880).
5. Farrar, R. A. H. *Antiquity* **53**, 211–216 (1979).
6. Sieveking, A. *Antiquity* **55**, 219–220 (1981).
7. d'Errico, F. *L'art Gravé Azilien* (CNRS, Paris, 1994).
8. Hedges, R. E. M., Law, I. A., Bronk, C. R. & Housley, R. A. *Archaeometry* **31**, 99–113 (1989).
9. Hedges, R. E. M. et al. *Radiocarbon* **34**, 306–311 (1992).
10. Bronk Ramsey, C. *Archaeol. Comput. Newsl.* **41**, 11–16 (1994).
11. Molleson, T. in *Trace Metals and Fluoride in Bones and Teeth* (eds Priest, N. D. & Van De Vyver, F. L.) 342–365 (CRC, Boston, MA, 1990).



Engraved surface and obverse (showing dating sample location) of the Sherborne bone. Photographs by F. d'Errico.

# Echoes of chaos

Karl Sigmund

**Frontiers of Complexity: The Search for Order in a Chaotic World.** By Peter Coveney and Roger Highfield. Faber/Fawcett: 1995. Pp. 462. £18.99, \$27.50.

**Hidden Order: How Adaptation Builds Complexity.** By John H. Holland. Addison-Wesley: 1995. Pp. 185. \$24 (hbk); £18.95 (pbk).

It is the way of fashion to go out of fashion. The new field of complexity seems to experience the same fate as its forerunner, chaos. After a few years of hope and hype, and a remarkable success on the public relations front, it is hitting lean times. A *Scientific American* suddenly turned sceptical appears to herald the decline by asking bluntly whether complexity is a sham.

No two researchers in the field agree on what defines complexity: but all make use of the same handful of examples: the brain, the Brazilian rainforest, the immune system and New York City. These systems are wonderfully intricate, evolve in time and offer daunting challenges for computer simulations. But can they be subjected to a unified approach? A battle-cry of chaos was that simple rules can lead to complicated behaviour: now it seems that complicated rules can also lead to complicated behaviour. Is that all it needs to keep pop science going?

Following in the wake of several bestsellers on the subject, *Frontiers of Complexity* and *Hidden Order* are latecomers and may have to face a recession. On the other hand, both sides of the complexity debate will find them useful for ammunition. The two books are as different as can be. *Frontiers of Complexity* is science journalism at its best, succinctly covering an impressively wide range of recent results and carefully tracing back their roots (mostly to Alan Turing and to John von Neumann). *Hidden Order*, by contrast, offers a highly personal manifesto of complexity, in the form of a series of easygoing lectures, by one of the godfathers of the field.

The authors of *Frontiers of Complexity* (a physicist and a science journalist who formerly collaborated on the well-received *Arrow of Time*) have produced an impeccably researched, amazingly up-to-date, crisply written and well-illustrated survey of a plethora of scientific topics that have recently grabbed the media's attention (neural networks, test-tube evolution and quantum computers, to name but a few). These mostly happen to be complex systems, but there is little to suggest an underlying unified science. Actually, apart from occasional sallies against misguided reductionism, the authors do not belabour this point. Extensive coverage is their aim. They have obtained interviews with all the right people, checked and double-checked their information and come up with more

than 60 pages of endnotes. The only (minor) weakness of their book is that it is so crammed full with information that it reads like 250 *Nature* News and Views articles back to back. The authors must have worked hard to come close, in every field, to the shortest possible description (whose size, incidentally, is one measure of complexity), but a reader relentlessly fed on such a diet of compressed information will



Emerging order — pattern 'bred' by Karl Sims through natural selection in a supercomputer.

occasionally long for some redundancy.

*Hidden Order* is much more relaxed, a casual meandering through vistas of computational ecologies. John Holland, the author, is one of the most durable figureheads of complexity. The first person to obtain a PhD in computer science, he has pioneered since the early 1960s the use of genetic algorithms, where populations of computer programs evolve under mutation, selection and recombination to achieve well-adapted solutions to complex problems. Holland's approach has proved its worth and spawned a vast field of parallel problem-solving techniques that aim at rivaling classical optimization methods.

In *Hidden Order*, Holland describes his more recent work on *cas* (complex adaptive systems). He unabashedly believes that there are general principles ruling all these examples, from cell to megalopolis; warns us — rather engagingly — that "the task of formulating theory for *cas* is more than usually difficult"; and then proceeds to sketch the construction of Echo — ultimately the mother of all models, but at this

stage a growing class of models in progress, intended to "rephrase" all instances of *cas* and to illustrate the emergence of complex structures through natural selection. Echo consists of populations of active elements called 'agents' that can reproduce in cyberspace. The rules are judiciously inspired from all kinds of *cas*, be they ecological (agents have to collect resources before they reproduce), genetical (the agents' capabilities are determined by strings), immunological (agents have tags reminiscent of epitopes), economical (agents can exchange resources) and so on. Given the right sort of coaxing, some agents ought eventually to aggregate and build hierarchical structures. The final stage should be a kind of supermodel able to mimic the embryogenesis of multicellular organisms just as well as the origins of multi-agent organizations such as industrial enter-

prises. There is, Holland admits, still a long way to go.

Will this type of work ever lead to a science of complexity able to help those tackling real complex systems, for instance embryologists? Hardly so. But that is not the point. In *Frontiers of Complexity* we read that the young Holland used to file his ideas on adaptation in folders labelled "Glass Bead Game". This is more than just an echo of the 1960s' infatuation with Hermann Hesse's novels. It pithily resumes the role of what erroneously is termed the 'science' of complexity: it is not a science in the usual sense at all, but, like the fictitious Glass Bead Game, an intellectually supercharged exercise aimed at relating and reflecting themes from diverse disciplines, and a way of organizing the rush of new information and of gaining familiarity with emerging insights. This is why science writers adore it. □

Karl Sigmund is at the Institut für Mathematik, Universität Wien, Strudlhofgasse 4, A-1090 Wien, Austria.

# Ditching causality

Peter Holland

## Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony.

By James T. Cushing. *University of Chicago Press*: 1994. Pp. 317. \$65, £51 (hbk); \$27, £21.50 (pbk).

**The Mystery of the Quantum World, second edition.** By Euan Squires. *Institute of Physics Publishing*: 1994. Pp. 191. £45, \$90 (hbk); £12.95, \$24.50 (pbk).

**An Interpretive Introduction to Quantum Field Theory.** By Paul Teller. *Princeton University Press*: 1995. Pp. 176. \$35, £23.50.

THE foundations of quantum mechanics as an independent subject provides a salutary lesson for those who see scientific development as driven by internal dynamics divorced from the social context of scientists. For what we today take for granted as a respectable and thriving field was, not so long ago, a dangerous move for a career-minded scientist. Its existence is due in no small measure to the struggles of a handful of dissidents, including some of this century's finest scientists (Einstein, de Broglie, Schrödinger), who wanted physics to be based on a clear and consistent ontology. They believed, as John Bell put it, "that vagueness, subjectivity and indeterminism are not forced on us by experimental facts, but by deliberate theoretical choice". They raised questions that the orthodoxy in theoretical physics, in place by the late 1920s, did not so much answer as ignore. As pointed out in James Cushing's fine book, one reason the dissenters were unable to establish their alternative was that they were divided in their views of what a desirable ontology should be, and so could not mount a coherent challenge to a united opposition.

The subject of Cushing's book is how a point of view that denied the possibility of causal space-time representations of atomic phenomena, the Copenhagen interpretation developed principally by Bohr, gained ascendancy over equally viable (in the sense of empirical adequacy) descriptions that did employ such representations. He recounts the events that led first to the crushing of de Broglie's tentative proposals for a space-time model in 1927 and then to the marginalization of David Bohm's fully fledged theory in 1952. Cushing argues persuasively that an historical error has been made in eschewing causal models and illustrates his case by comparing the quality of the explanations provided by the two world views for typical effects such as two-slit interference. His thesis is that Copenhagen triumphed not because it had more compelling technical arguments but because it got in first, became entrenched and used the dead-weight of

authority to see off all competitors.

It was a fortuitous victory. Cushing speculates how the story might have turned out had the order of certain ("historically contingent") events been different. Historians may see this as an idle exercise but the author's aim is to show that scientific arguments alone did not justify the wholesale ditching of causality as a goal of physics. This decision, whatever its merits, did not flow from facts forced on us by 'nature' but from extra-scientific pressures. Along the way, Cushing reveals important but little known material, such as an unpublished hidden-variable interpretation of Einstein.

I am not sure Cushing has got to the bottom of this story, for Bohm's ideas have been anathema even to physicists who have sustained a critical attitude in questions of interpretation. Renowned scientists have spent years promoting the bizarre suggestion that the Universe 'splits' whenever a choice is available for the future behaviour of an electron, in preference to Bohm's simple notion that an electron moves along a track in space. Another important cultural issue not addressed is why it should be now, after 40 years of neglect, that many workers are attracted to Bohm's theory. But the author is to be applauded for recognizing the need for this analysis and for producing a sustained and well written account that will undoubtedly stimulate others to look at other aspects of a remarkable story, especially at the crucial role of dissidence in scientific culture.

Euan Squires's well known book is one of the few that successfully straddle the huge gulf between formal texts and traditional popularizations of quantum mechanics. It is aimed at the general reader but I suspect its natural place is as an adjunct to an undergraduate course. In that context it will play a valuable role in showing students that the physical interpretation of the techniques they are learning is a matter of continuing and unresolved debate. The author gives clear descriptions of the problem of quantum reality, the role of consciousness, and non-locality, including an accessible account of Bell's theorem. There are helpful anecdotes, including an account of his own experience as a student in the 1950s when discussion of these issues was frowned upon. Squires brings the first edition up to date through an additional chapter on recent developments in collapse theories, the Bohm model and Lorentz invariance. It is a pity it was not possible to integrate this material into the main text, for Squires's own views on these topics, to which he himself has made important contributions, have moved on since the previous edition.

It is fair to say that our comprehension of quantum field theory is more slender even than that of elementary quantum mechanics. Paul Teller faces up to this issue by analysing the meaning of the terms used in traditional textbook treatments of field

theory. (He unfortunately does not discuss the currently popular functional techniques that are pregnant with possibilities for imaginative new interpretations.) What he does do is examine how the concepts of 'particle' and 'field' are to be understood and asks whether they can be smoothly connected. He discusses how the notion of an 'operator-valued field' is related to our customary concept of a field and shows, correctly I think, that a direct association of the two is misleading. Although questions may be raised about the physical picture implied by his use of the 'propensity' concept, I believe this is a valuable book. Indeed, the specific interpretative proposals he makes could have an influence on hidden-variable theories. The author expresses scepticism about hidden-variable theories, but if he had adopted a definition more in accord with their actual characteristics he might have found fruitful overlap between the two programmes. At the risk of labouring the point, Bohm's theory has many interesting things to say about space-time pictures of quantum fields, and it would be interesting to explore how it impinges on Teller's considerations.

These three books reflect a welcome trend towards a reassessment of the notions of meaning and visualization in contemporary science. Scientists from other disciplines would be surprised to learn how much effort has been expended in expunging causal imagery from physics. Yet for all their protests to the contrary, even the most committed Copenhagen supporter privately employs images of the micro-domain because these are indispensable when applying quantum mechanics in unfamiliar situations. It is a merit of Bohm and others that they were honest about this, and sought to go beyond the odd amalgam of classical ideas espoused by Bohr to find a more satisfactory iconography. □

Peter Holland is at the School of Interdisciplinary Sciences, University of the West of England, Bristol BS16 1QY, UK.

## Corrections

In Peter Knight's review of *Optical Coherence and Quantum Optics* by Leonard Mandel and Emil Wolf (*Nature* **378**, 348; 1995), a line was inadvertently missed out in the editing, leading to the incorrect statement that Wolf is an experimentalist. The fourth and fifth sentences of the second paragraph should in fact read: "Wolf was the originator of many of our fundamental ideas of coherence, and with Mandel he pioneered much of our understanding of photon statistics. Mandel was the founder of the experimental study of nonclassical light and the first to generate laboratory sources of light of totally quantum origin such as anti-bunched light and the correlated twin beams of light from parametric amplifiers."

In Peter J. Bowler's review of *Gregor Mendel* by Vitezslav Orel (*Nature* **378**, 100; 1995), Brno is described as being in Austria rather than in the Czech Republic. Both mistakes were made in the editorial office of *Nature*. Our apologies.



# What thoughts are made of

Stevan Harnad

**The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain.** By Paul M. Churchland. MIT Press: 1995. Pp. 326. \$29.95, £19.95.

**Journey to the Centers of the Mind: Towards a Science of Consciousness.** By Susan A. Greenfield. W. H. Freeman: 1995. Pp. 221. \$24.95, £17.95.

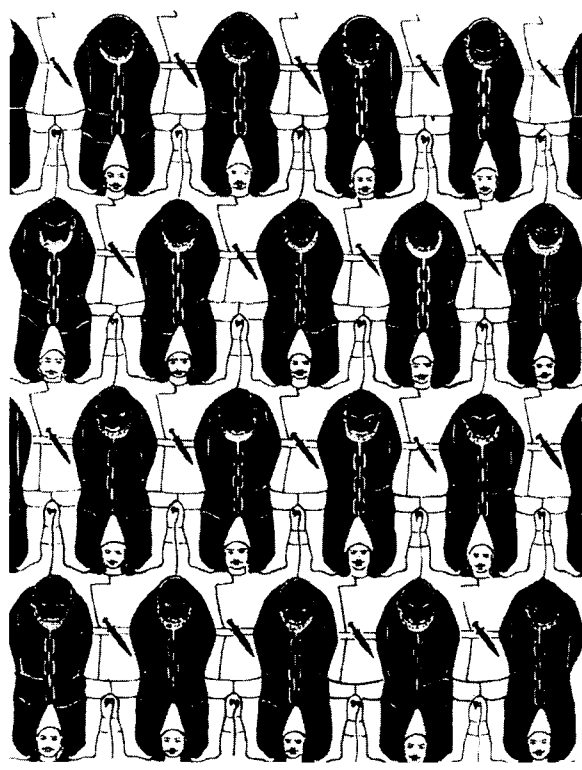
UPON minor prodding by Hamlet, Polonius shows himself quite capable of seeing a cloud as, successively, a camel, a weasel or a whale. Let us call this remarkable capacity to see one thing in terms of another — and the pattern-completing cognitive closure, hence insight, that it brings with it — the human hermeneutic capacity. It is responsible for everything from our ability to answer wrenching existential questions with a Bible and a pin to our inability to avoid seeing every dog as looking like its owner. But when it comes to explaining the mind, it is unclear whether this Polonian predilection of ours is a boon or a bane.

Paul Churchland is a philosopher of what used to be called the 'eliminative materialist' turn of mind. The objective was to eliminate our everyday view of the mind in favour of a new, scientific one, in which there is no longer a mind-body problem — the difficulty we all seem to have in seeing how mental states (feelings, thoughts, indeed any form of experience) can be physical, or even objectively explicable. It is fashionable these days to blame the problem on Descartes, but the ancient pedigree of religion and belief in the immaterial soul hints that it may be more deeply rooted than anything planted in our minds in the seventeenth century.

Churchland makes no secret of his strategy to uproot it: "I hope to make available here a conceptual framework of sufficient richness and integrity that you will be able to reconceive at least some of your own mental life in explicitly neuro-computational terms". And so he does, surveying a good bit of contemporary cognitive neuroscience, including brain and behavioural modelling (to which he himself has made some original computational contributions, putting into practice his principle that philosophical problems can be solved by assimilating them to science).

Churchland sets great store by 'neural nets'. The term is ambiguous, because on the one hand there are undeniable networks of neurons in the brain and on the other there are also networks of artificial or notional 'neurons' implemented on computers. Worse still, the nets implemented on computers are usually discrete,

PATTERN generations — "Unity" by Khudu Mamedov, an Azerbaijani crystallographer. The space group of this periodic drawing is  $p1$ , with a basic motif of an old and a young man. The repetition of the uniform shapes truly satisfies the two-dimensional space group, although closer inspection reveals distinct individual facial expressions, especially of the old men. Taken from the second edition of *Symmetry through the Eyes of a Chemist* by István Hargittai and Magdolna Hargittai. "Education by aesthetic appeal", said *Nature's* reviewer of the first edition. Plenum, \$102.



serial simulations of what are intended to be continuous, parallel, distributed systems. Churchland is betting on the first and the third of these — indeed he takes them to be one and the same thing; but he is an explicit critic of purely computational models of the mind, in this respect making common cause with other critics of computation such as the philosopher John Searle and the mathematician Roger Penrose, although he is at pains to dissociate his own arguments from theirs, which he finds wanting.

Churchland first introduces the notion of 'vector activations' by surveying the sensory physiology of taste, colour and smell. Each of these senses has detectors, with complex sensations involving the activity of combinations of them. Colour is a familiar example. Among the cones in the retina, there are some that are selectively activated by red, others by green and still others by blue. Leaving out a few complicating details, it can be said that linear combinations of the activation levels of each of these three detectors generate all the colours we are capable of seeing. So it seems correct to describe colour space vectorially as a sensory activation space.

As a similar story can be told about the other senses, it seems reasonable to think of all sensory input as vector activation of some kind. Indeed, vector activation appears to be a very general notion, because for humans all 'input' is sensory input. Churchland next moves to a level higher than raw sensory processing to the more complex case of face recognition. In

doing so, he also takes leave of brain and behavioural evidence and passes to computational evidence: like ourselves, artificial neural nets turn out to be able to classify faces (both facial expressions and facial identities), and they do so by partitioning vector activation space. Neural nets consist of layers: the input layer is the sensory input vector; the output layer is the motor activation vector; and various hidden layers in between are internal vectors, their components interconnected with the input and output vectors. The net accomplishes what it does by adjusting the strength of its interconnections, and hence the strength of the activations of the input, hidden and output vectors, on the basis of various built-in learning rules.

Two suggestive concepts emerge for Churchland from the computational modelling work: those of prototypes (the centres of regions in the partitioned vector space) and 'vector completion' (input vectors that are correctly classified even though they are incomplete). Mere feedforward nets (in which activation flows in only one direction, from input to hidden layer to output) are contrasted with recurrent nets in which there are also descending feedback connections and in which repetitive activation cycles are possible. Feedforward nets seem to be data-driven learners of static invariants, as in face classification, whereas recurrent nets can learn temporal sequences, such as whether certain symbol strings are grammatical or ungrammatical (although I think Churchland misconstrues this particular toy model of a tiny fragment of

grammar, taking it to be a refutation of Chomsky's evidence and arguments for an unlearnable, hence inborn, universal grammar).

With this repertoire of concepts, we learn that "[i]f having a feedforward neural architecture is what allows one to discriminate instances of prototypical things, then having a recurrent neural architecture is what provides one with the further capacity to discriminate instances of processes" and that "[w]ithout [the vectorial sequences generated within a well-trained recurrent network]... we would have no concept of temporal extension or of causal processes at all."

This is eliminative materialism at work. Chapter 5 is a grander and grander series of hermeneutic exercises (we are here playing Polonius to Churchland's Hamlet) on the capacity of recurrent nets to recognize the past, understand causality, disambiguate figures and make scientific discoveries. The trouble is that in place of empirical evidence that the actual behavioural capacities of nets can scale all the way up to our own in this way, the few tiny toy demonstrations that exist are instead ratcheted up hermeneutically, by being subjected to a Protean mentalistic interpretation in which recurrent nets turn out to have the seven core properties of consciousness singled out by Churchland: short-term memory, independence of sensory input, steerable attention, alternative interpretation capability, absence in sleep, presence in dreaming and unity of experience. And conscious knowledge turns out to be what passes along 'auto-connected' pathways (directly connected to the information source, as in my knowledge that my own bladder is full) in contrast to 'hetero-connected' pathways (as in my knowledge that *your* bladder is full). (The reason this vocabulary doesn't quite do the eliminative trick for me is that even my hetero-connected knowledge that your bladder is full strikes me as conscious; moreover, the auto-connected thermostat strikes me as being unlikely to be conscious of heat. Ditto for any servosystem or recurrent net.)

It is not that it is impossible that this is what the mind amounts to; it just seems grossly premature to say so on the evidence to date. Most of the mileage seems to be coming instead from our Polonian impressionability and credulity (and I take this to be bad news for the eliminativist programme as a whole, because it shows that we can easily be brainwashed into thinking that the mind-body problem has been solved when it hasn't).

I also think Churchland underestimates the power and purpose of the Turing test, dismissing it as the trivial game to which the Loebner prize (offered for the computer program that can fool judges into thinking it is human) has reduced it, whereas in reality it is an exacting empiri-

cal criterion: the candidate model for the mind must have our full behavioural capacities — so fully that it is indistinguishable from any of us, to any of us (not just for one contest night, but for a lifetime). Scaling up to such a model is (or ought to be) the programme of that branch of reverse bioengineering called cognitive science. It is harmless enough to do the hermeneutics after the research has been successfully completed, but self-deluding and question-begging to do it before.

Following her own informative and well written survey of the neurobiological data, Susan Greenfield, in *Journey to the Centers of the Mind*, proceeds straight to the hermeneutics without even pausing over the problem of performance modelling: she offers a consciousness criteria list even shorter than Churchland's, consisting of 'concentric' epicycles — mind patterns of which she spies an abundance nestled in the brain clouds. □

Stevan Harnad is in the Department of Psychology, University of Southampton, Highfield, Southampton SO17 1BJ, UK.

## Broad perspectives

Janette Atkinson

**Visual Development.** By Nigel W. Daw. Plenum: 1995. Pp. 228. \$54, £36.

THE past 30 years has seen enormous advances in our understanding of visual development and deprivation. Nigel Daw's excellent new book is a celebration of the success of research in this area, and is the first text to unify research results and ideas from behavioural, clinical, anatomical and physiological approaches. This unique feature makes this a book that can be read with benefit by students in psychology, neurobiology, ophthalmology, paediatric neurology and optometry, and its useful glossary makes it accessible even to beginners in the field of visual research. The book also provides a good summary for more advanced researchers likely to have expertise and up-to-date knowledge in some areas but who lack the time to keep up with advances right across the field; it provides a readable, well-illustrated and succinct account to fill the gaps in their knowledge.

Daw begins with a brief historical introduction to clinical problems of visual deprivation, from the seventeenth-century debates of Molyneux and Locke to the work of the Nobel prizewinners David Hubel and Torsten Wiesel, the starting point of the modern research that makes up the bulk of the book. Daw points out that Hubel and Wiesel's work allowed the first clear correlations to be

made between clinical observations on strabismus (abnormal alignment of one or both eyes) and deprivation amblyopia (impaired vision with no discernible damage to the eye or optic nerve) and measurable changes in the central visual system.

There then follows a 20-page summary of the functional organization of the visual system, from the retina to cortical streams, which provides the necessary background for an understanding of the three main sections of the book. In the first section, Daw considers normal visual development from three perspectives, psychophysics, anatomy and physiology, and examines the links between them. New techniques and results are succinctly summarized and illustrated around a discussion of which components of development depend on sensory inputs and which are hard-wired — a distinction essential for a proper understanding of the subject of the second section, visual deprivation. Here Daw provides a detailed discussion of the neural correlates of strabismus and amblyopia. He ends this section by addresses the thorny question of human critical periods and differing plasticity in different subsystems, an area of considerable clinical importance for devising sensible medical guidelines for effective intervention and treatment of strabismus and amblyopia. The third section covers more recent research on the mechanisms of plasticity. Daw takes long-term potentiation as a model of the way in which the plastic changes in development may be brought about at the cellular and molecular level, and considers NMDA glutamate receptors to be the prime candidates for mediating early learning.

In a stimulating concluding chapter on future studies and unanswered questions, the author outlines the present controversies and debates, such as the relative roles of excitatory and inhibitory influences in shaping changes in ocular dominance. He also draws attention to the growing field of developmental neurobiology as applied to more complex visual processing. Here there are unanswered questions on the mechanisms controlling the plasticity and development in visuomotor control, spatial cognition, object recognition and integration between analyses in different cortical streams. Let us hope that in a few years, when, hopefully, some of these questions have been answered, someone will cover these areas in a book as clearly written as Daw's.

In the meantime, Daw has given us a lucid and readable exposition, covering in depth and breadth an important era of vision research. □

Janette Atkinson is at the Visual Development Unit, Department of Psychology, University College London, Gower Street, London WC1E 6BT, UK.

# The structure of an intermediate in class II MHC maturation: CLIP bound to HLA-DR3

Partho Ghosh<sup>\*</sup>, Miguel Amaya<sup>†</sup>, Elizabeth Mellins<sup>†</sup> & Don C. Wiley<sup>\*‡§</sup>

<sup>\*</sup> Department of Molecular and Cellular Biology, and <sup>‡</sup> Howard Hughes Medical Institute, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

<sup>†</sup> Children's Hospital of Philadelphia and Children's Seashore House, Philadelphia, Pennsylvania 19104, USA

**A complex between HLA-DR3 and a fragment of invariant chain called CLIP was isolated from a human cell line defective in antigen presentation and its X-ray crystal structure determined. Previous data indicate that this complex is an intermediate in class II histocompatibility maturation, occurring between invariant chain-DR3 and antigenic peptide-DR3 complexes. The structure shows that the CLIP fragment binds to DR3 in a way almost identical to that in which antigenic peptides bind class II histocompatibility glycoproteins. The structure is the substrate for the loading of antigenic peptides by an exchange process catalysed by DM.**

CLASS II histocompatibility membrane glycoproteins ( $\alpha\beta$ ) bind antigenic peptides in endosomal compartments of antigen-presenting cells for 'presentation' at the cell surface to T cells. A second membrane glycoprotein, invariant chain (Ii), complexes with  $\alpha\beta$  in the endoplasmic reticulum to stabilize it in the absence of a bound peptide and to target it to the endocytic pathway (reviewed in ref. 1). Ii, which blocks the binding of antigenic peptide to  $\alpha\beta$ , is removed from  $\alpha\beta$  by stepwise proteolysis in endosomes before antigenic peptide loading. In several cells defective in antigen presentation, complexes of  $\alpha\beta$  with a nested set of 20–24 residue Ii fragments (within residues 81–104) called CLIP (class II associated invariant chain peptide) accumulate at the cell surface<sup>2–5</sup>, raising the possibility that  $\alpha\beta$ -CLIP is an intermediate in class II molecule maturation<sup>2,4–6</sup>. The observations that *in vitro* proteolysis of  $\alpha\beta$ Ii isolated from cells generates  $\alpha\beta$ -CLIP, and that pulse-chase experiments in cells demonstrate the transient formation of  $\alpha\beta$ -CLIP, are further evidence that  $\alpha\beta$ -CLIP is an intermediate that is generated naturally during class II maturation<sup>6,7</sup>.

The CLIP segment of Ii has been implicated in several of the activities of Ii. Genetic truncation and deletion experiments<sup>8,9</sup> demonstrate that the CLIP segment is necessary for the *in vivo* activities that promote  $\alpha\beta$  assembly and stabilize  $\alpha\beta$  dimers against permanent aggregation with chaperones in the endoplasmic reticulum. The CLIP segment is also necessary for the *in vitro* Ii activity of inhibiting peptide binding to  $\alpha\beta$ <sup>9,10</sup>. Differing affinities of CLIP for different alleles of DR and analysis of the binding of substituted CLIPs<sup>6,11–13</sup> provide compelling evidence that CLIP binds in the  $\alpha\beta$  peptide binding site. However, alternative allosteric models have been proposed<sup>8</sup>, and studies of a mutant class II molecule<sup>14</sup> and the inhibition motif of a short (92–104) fragment of CLIP<sup>15</sup> have been interpreted to favour an alternative mode of binding.

Various cell lines deficient in antigen presentation have been shown to lack HLA-DM<sup>16,17</sup>, which is an MHC-encoded protein. The accumulation of DR3-CLIP on the surface of these cells suggests that DM is required to remove CLIP from DR3 in endosomes so that antigenic peptides are able to bind<sup>16,17</sup>. DM has recently been shown to stimulate the release of CLIP from class II MHC *in vitro*<sup>18–20</sup>. The peptide release effect of DM is limited to CLIP and to at least one other peptide<sup>18</sup>, and does not extend to all peptides.

We have purified DR3-CLIP from a cell line deficient in antigen presentation<sup>21</sup> that lacks DM<sup>16,17</sup>, and determined its struc-

ture from crystals grown at the pH (4.5) of endosomes. The structure reveals that CLIP occupies the peptide binding site and binds in a way almost identical to the way in which a peptide antigen, influenza virus haemagglutinin (HA 306–318), binds to DR1 (ref. 22). The contacts between DR3 and CLIP suggest that the CLIP segment of Ii would stabilize most class II molecules almost as much as an antigenic peptide. We measured the half-life of CLIP bound to DR3 (at 37 °C, pH 4.5) to be almost 4 hours, long enough to require an exchange process, such as that mediated by DM, to remove CLIP from most subtypes of class II MHC before antigenic peptides can bind *in vivo*. The conformation of DR3 in the DR3-CLIP complex is only slightly different from that of DR1 in DR1-HA, although it is sufficient to explain antigenic differences without recourse to allosteric transitions or novel conformations. We show that the structure observed here is the substrate for the DM exchanger, limiting the possible mechanisms for the catalytic activity of DM.

## CLIP occupies the peptide binding site

The X-ray crystal structure of DR3-CLIP was determined by molecular replacement. Details of the structure determination (Fig. 1a) and refinement are presented in Table 1.

The peptide binding site of DR3 is filled with a continuous segment of electron density, indicating that CLIP binds in the site normally occupied by peptide antigens<sup>22–25</sup> (Fig. 1b). CLIP does not appear to bind to other sites on DR3, as no substantial electron density is observed that is not due to DR3, carbohydrate or water. The excellent quality of the peptide electron density shows that the DR3 is occupied almost exclusively by CLIP. Of the 24 residues in CLIP (L81–M104), 15 are visualized in the binding site (P87–A101), whereas 4–5 residues at the amino terminus and 2–3 residues at the carboxy terminus extend out of the site and are presumably disordered in the crystal (Fig. 1b; Table 2).

The composition, stability and ability for peptide exchange of DR3-CLIP, which was derived from DM-deficient cells and solubilized with papain, are comparable to those of DR3-CLIP complexes reported previously<sup>2–5</sup>. Three peptides have been shown by mass spectroscopy to predominate in both DR3 recovered from crystals and in DR3 solubilized by papain digestion before crystallization: P82–M104, P82–P103, and K83–P103 (Fig. 1b; Table 2). They are among the major forms of CLIP eluted from DR3 on the cell line deficient in antigen presentation<sup>4,5</sup>. Only small quantities of peptides with the first amino acid of CLIP (81L) were observed after papain

§ To whom correspondence should be addressed.

TABLE 1 Structural analysis and refinement

| Crystallographic data         |         | Refinement results |        |
|-------------------------------|---------|--------------------|--------|
| Resolution (Å)                | 15–2.75 | 3–2.75             | 6–2.75 |
| Independent reflections       | 13,021  | 2,876              | 11,772 |
| Multiplicity                  | 3.2     | 3.0                | 3,095  |
| Completeness (%)              | 96.9    | 97.2               | 32.5%  |
| Average $I/\sigma(I)$         | 6.7     | 2.3                | 24.6%  |
| $R_{\text{merge}}$ (%)        | 7.1     | 31.4               |        |
| Refinement results            |         |                    |        |
| Reflections                   |         |                    |        |
| Atoms                         |         |                    |        |
| $R_{\text{free}}$ †           |         |                    |        |
| $R_{\text{crist}}$ †          |         |                    |        |
| r.m.s. deviations             |         |                    |        |
| Bonds (Å)                     |         |                    |        |
| Angles (deg) ‡                |         |                    |        |
| Dihedrals (deg) ‡             |         |                    |        |
| Improper (deg) ‡              |         |                    |        |
| B-factors (Å <sup>2</sup> ) § |         |                    |        |
| Real space fit                |         |                    |        |
| Average B-factors             |         |                    |        |
| HLA-DR3 (Å <sup>2</sup> )     |         |                    |        |
| CLIP (Å <sup>2</sup> )        |         |                    |        |

**Purification and crystallization.** HLA-DR3 was purified<sup>4</sup> from 9.5.3 cells<sup>21</sup>, solubilized in  $\beta$ -octylglucoside, and cleaved with papain<sup>40</sup>; it was further purified using a BioSil S3000 column and concentrated to  $\sim 15 \text{ mg ml}^{-1}$ . Crystals were grown using conditions similar to those reported for other DR molecules<sup>21</sup>: 12% PEG 4K, 100 mM MgCl<sub>2</sub>, 100 mM acetate buffer, pH 4.5. Crystals belong to space group  $P3_212$  with cell dimensions  $a = b = 78.5 \text{ Å}$ ,  $c = 159.1 \text{ Å}$ . **Data collection and processing.** Diffraction was collected as oscillations of  $1^\circ$  (CHESS F-1;  $\lambda = 0.91 \text{ Å}$ ) on Fuji phosphor plates from two flash-frozen crystals (100 K); before freezing, crystals were soaked for 5 min in 15% glycerol, 15% ethylene glycol, 16% PEG 4K, 100 mM MgCl<sub>2</sub>, 100 mM acetate, pH 4.5; mosaicity of frozen crystals was  $\sim 0.2^\circ$ . Data were indexed and integrated using DENZO (Z. Otwinowski, personal communication), and scaled and reduced using ROTAVATA/AGROVATA, POSTREF and TRUNCATE<sup>42</sup>. Partial reflections were summed. **Molecular replacement.** A molecular replacement solution was identified with Amore<sup>43</sup> using a single  $\alpha\beta$  DR1 (ref. 22) heterodimer, for which 19 residues differing between DR1 and DR3 were substituted by Ala, as the model. The rotation (15–3 Å) and translation (8–4 Å) solutions were the highest peaks in their respective maps. After rigid-body refinement (8–3.2 Å, with Amore), the correlation coefficient (c.c.) of the rotation and translation solution was 0.59 and the  $R_{\text{crist}}$  42.7% (next-highest solution: c.c., 0.27;  $R_{\text{crist}}$  54.8%). The translation search identified  $P3_212$  as the correct enantiomorph. The asymmetric unit of the crystal contains a single  $\alpha\beta$  heterodimer of DR3, and the same dimer of  $\alpha\beta$  heterodimers observed previously<sup>22–25</sup> was observed here as a crystallographic dimer. **Model building and refinement.** A random 10% of reflections were omitted from refinement for  $R_{\text{free}}$  calculation. Before refinement, the model B-factor was set to  $45 \text{ Å}^2$  (initial  $R_{\text{free}}$ , 44.1%;  $R_{\text{crist}}$ , 45.7%, 10–3.5 Å). The model was rigid-body refined using X-PLOR<sup>44</sup>, and residues specific to DR3 were built into  $2F_o - F_c$  omit maps (15–3.5 Å) using O<sup>45</sup>. Preferred rotamer conformations were used when possible, and a database of refined structures was used for main-chain rebuilding<sup>45</sup>. Only procedures which minimized  $R_{\text{free}}$  were followed; the difference between  $R_{\text{free}}$  and  $R_{\text{crist}}$  was kept as small as possible. After a single rebuilding and refinement step (positional and group B-factor, 6–3.5 Å), an anisotropic B-factor tensor ( $B_{11} = B_{12} = B_{22} = -10 \text{ Å}^2$ , and  $B_{33} = 10 \text{ Å}^2$ ) determined from a Wilson plot was applied, resulting in a 1% drop in  $R_{\text{free}}$ . The resolution limit was increased gradually to 2.75 Å through subsequent rebuilding and refinement cycles, which included positional and restrained atomic B-factor refinement, geometric regularization, and occasionally simulated annealing. The model was rebuilt into weighted (with SIGMA<sup>42</sup>)  $2F_o - F_c$  omit maps (15–2.75 Å), in which either approximately one-tenth of the model or entire protein domains were omitted, or into simulated annealing omit maps. The peptide was modelled into  $F_o - F_c$  density after 6 cycles of rebuilding and refinement ( $R_{\text{free}}$ , 34.8%;  $R_{\text{crist}}$ , 27.8%; 6–2.75 Å). After 10 cycles ( $R_{\text{free}}$ , 33.4%;  $R_{\text{crist}}$ , 23.8%; 6–2.75 Å), waters within 2.6–3.4 Å of a hydrogen bonding donor or acceptor were built into  $\geq 3\sigma F_o - F_c$  density. Positionally unstable waters or those with refined B-factors greater than  $60 \text{ Å}^2$  were removed. The final model contains 22 waters. Only one N-acetylglucosamine monosaccharide unit could be fit at the N-linked glycosylation sites,  $\alpha 78\text{N}$  and  $\alpha 118\text{N}$ . No carbohydrate could be modelled at the third N-linked glycosylation site,  $\beta 19\text{N}$ . All residues have allowed  $\phi$ ,  $\psi$  angles, and coordinate error is estimated to be  $0.35\text{--}0.45 \text{ Å}$ <sup>46</sup>.  $2F_o - F_c$  maps show clear electron density for residues 5–180 of the  $\alpha$ -subunit and residues 5–191 of the  $\beta$ -subunit. Three small breaks in the main-chain density occur in loop regions of the  $\beta_2$  domain (at  $\beta 109$ ,  $\beta 172$  and  $\beta 189$ ). The side chains of 15 residues on surface loops are disordered ( $\alpha 123\text{R}$ ,  $\beta 19\text{N}$ ,  $\beta 22\text{E}$ ,  $\beta 23\text{R}$ ,  $\beta 105\text{K}$ ,  $\beta 107\text{Q}$ ,  $\beta 109\text{L}$ ,  $\beta 110\text{Q}$ ,  $\beta 111\text{H}$ ,  $\beta 136\text{Q}$ ,  $\beta 138\text{E}$ ,  $\beta 166\text{R}$ ,  $\beta 167\text{S}$ ,  $\beta 189\text{R}$  and  $\beta 191\text{R}$ ). These residues are modelled, but atoms beyond C $\beta$  were omitted in the later refinement stages. A break occurs in the side-chain density of  $\beta 74\text{R}$  between C $\delta$  and C $\zeta$ . Although the guanidinium of  $\beta 74\text{R}$  is positioned in electron density, this density may instead correspond to a solvent molecule. The side-chain amide and carboxyl groups of Q100 of CLIP are not found in electron density, and therefore calculations of buried surface area do not include these two groups.

\*  $R_{\text{merge}} = (\sum_i \sum_j |I_{h,i} - I_{h,j}|) / (\sum_i \sum_j I_{h,i})$ , where  $I_h$  is the mean intensity of symmetry-related reflections,  $I_{h,i}$ .

†  $R_{\text{value}} = (\sum_{\text{hkl}} |F_{\text{obs}} - F_{\text{calc}}|) / \sum_{\text{hkl}} F_{\text{obs}}$ , where  $R_{\text{free}}$  is calculated for a randomly chosen 10% of reflections ( $F > 0$ ) omitted from refinement, and  $R_{\text{crist}}$  is calculated for the remaining 90% of reflections ( $F > 0$ ) included in refinement.

‡ Root-mean-square deviations from ideal values.

§ Root-mean-square deviations between B-factors of bonded atoms.

|| Real-space fit correlation coefficient<sup>45</sup> calculated from  $2F_o - F_c$  electron density and the refined model of DR3-CLIP.

TABLE 2 Nature and stability of DR3-CLIP isolated from DM-deficient cells

|                             |                  | Observed $m/z$ | Calculated $m/z$ |
|-----------------------------|------------------|----------------|------------------|
| Peptides isolated           | P82–M104         | 2563.41        | 2564.27          |
| from crystals of            | P82–P103         | 2432.15        | 2433.08          |
| DR3                         | K83–P103         | 2334.82        | 2335.96          |
| Conditions                  |                  |                |                  |
| Half-life of DR3-CLIP †     | 3.7 h            | pH 4.5, 37 °C  |                  |
| Effect of HLA-DM ‡          | Exchange of CLIP | pH 4.5, 37 °C  |                  |
| SDS-stability of DR3-CLIP § | Unstable         | 1% SDS, 25 °C  |                  |

\* Masses of peptides, which were acid-extracted<sup>35</sup> from papain-cleaved DR3 and from washed and solubilized crystals of DR3, were determined by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry<sup>36</sup>. The three peptides above were the predominant ones in mass analyses of pre- and postcrystallization samples. Because of PEG suppression in the postcrystallization samples, peptides that were of very low abundance in the precrystallization sample were not detectable in the postcrystallization sample. The shortest peptide detected was K83-L102, and a peptide beginning at the first residue of CLIP, L81-P103, was also detected, both in very low quantities and only in the precrystallization sample.

† Soluble DR3 at a concentration of  $2 \mu\text{M}$  was incubated in 100 mM acetate buffer, pH 4.5 at 37 °C for various times and monitored by absorbance,  $A_{280}$ , for quantity of DR3-CLIP complex remaining using high-performance liquid chromatography (HPLC) gel filtration (BioSil S3000); tyrosine was included as an internal standard. The initially homogeneous DR3-CLIP gel filtration peak yielded over time a separable mixture of aggregates and DR3-CLIP complexes; the aggregates are probably an indication of peptide dissociation and the formation of 'empty' molecules<sup>29</sup>. The half-life was determined from a first-order exponential fit of the data between 0 and 300 min; beyond 300 min, the amount of DR3-CLIP remains at a steady-state level. Detergents such as  $\beta$ -octylglucoside substantially decreased the DR3-CLIP half-life as previously reported<sup>6</sup>, as did Zwittergent (data not shown), which was used in studies reporting short half-lives of DR2-CLIP complexes<sup>15</sup>.

‡ Papain-solubilized DR3, from DM-deficient cells, at a concentration of  $20 \text{ nM}$  was incubated with or without  $4 \mu\text{M}$  soluble, recombinant HLA-DM<sup>18</sup> and with  $5 \mu\text{M}$  biotin-labelled IgCk (37–51) (KVQWKVDNALQSGNS) peptide for 4 h at pH 4.5, 37 °C (D. Zaller, personal communication). Incubation with DM led to a threefold increase in the amount of CLIP exchanged for IgCk peptide, similar to results reported for full-length DR3-CLIP (compare with Fig. 4a in ref. 18).

§ DR3-CLIP from samples used for crystallization and from solubilized crystals were unstable in SDS-PAGE in the absence of boiling, yielding monomers rather than heterodimers.

treatment (Table 2), but this should have no structural consequences because CLIP with the first two residues (81–82) missing has the same stability on DR<sup>26</sup> and participates in the DM exchange reaction in the same way as CLIP (81–104)<sup>19</sup>. Furthermore, the soluble DR3-CLIP complexes used for crystallization were shown to participate in the DM exchange reaction (D. Zaller, personal communication), similar to complexes purified without papain treatment from DM-deficient cells (Table 2).

We also observed that DR3-CLIP complexes after papain digestion and resolubilized from crystals have the characteristic instability to sodium dodecylsulphate (SDS) (Table 2) of DR3-CLIP reported previously<sup>2–5</sup>.

## Interactions between CLIP and DR3

The binding of CLIP to DR3 is remarkably similar to the binding of both the influenza virus HA peptide<sup>22</sup> and a collection of self-peptides<sup>23,24</sup> to DR1. Bound CLIP is extended in a polyproline type II conformation<sup>22</sup> (Fig. 2a); it forms a nearly identical, extensive network of hydrogen bonds to conserved class II MHC residues<sup>22</sup> (Fig. 2a); and its side chains extend into the



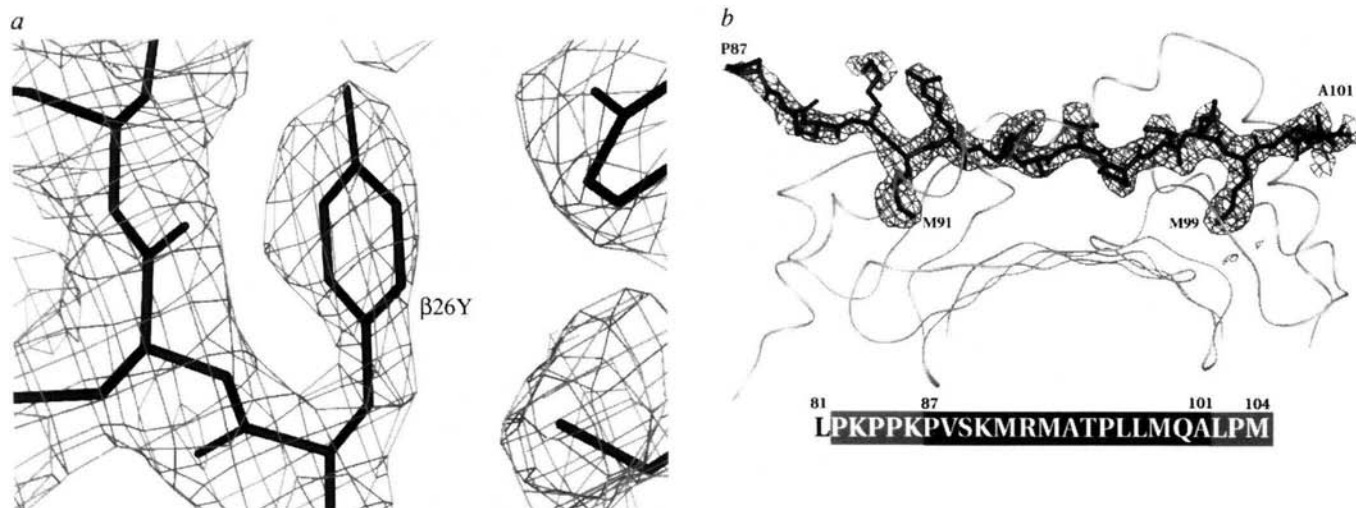


FIG. 1 Electron density for CLIP. *a*, Electron density for β26Y, a residue differing between DR3 and DR1 (β26L), is clearly present in a  $2F_o - F_c$  map (15–2.75 Å, contoured at  $1\sigma$ ) calculated before any rebuilding or refinement (except for rigid-body) of the molecular replacement model. In the model, β26 is replaced by Ala (see Table 1 legend). This figure and all subsequent figures, except for 2b, were made with MidasPlus<sup>47</sup>. *b*, Difference electron density shows that 15 residues of CLIP (P87–A101) are visible in the peptide binding site. Two methionines, M91 and M99, form the prominent density near the ends of CLIP and point down towards the floor of the binding site. The electron density is from

an omit  $F_o - F_c$  map (15–2.75 Å, contoured at  $2\sigma$ ) calculated from the refined model of DR3-CLIP from which CLIP was omitted. The  $\alpha_1\beta_1$  peptide-binding domain ( $\alpha 5$ – $\alpha 80$ ,  $\beta 5$ – $\beta 94$ ) of DR3 is shown in a wire trace with the  $\beta$ -sheet floor of the binding site at the bottom, the  $\alpha_1$  domain  $\alpha$ -helix in back, and the  $\beta_1$  domain  $\alpha$ -helix in front. Below the wire trace is the sequence of CLIP (L81–M104) with residues observed in the crystal structure shown in the central black box and disordered residues not observed in the structure shown in the surrounding grey boxes.

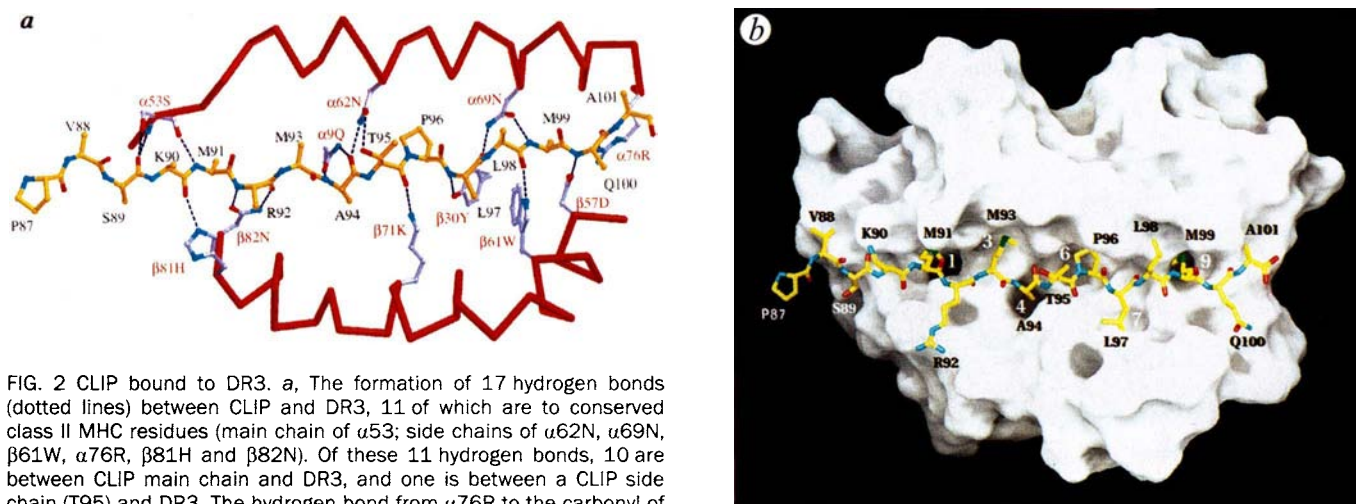


FIG. 2 CLIP bound to DR3. *a*, The formation of 17 hydrogen bonds (dotted lines) between CLIP and DR3, 11 of which are to conserved class II MHC residues (main chain of  $\alpha 53$ ; side chains of  $\alpha 62N$ ,  $\alpha 69N$ ,  $\beta 61W$ ,  $\alpha 76R$ ,  $\beta 81H$  and  $\beta 82N$ ). Of these 11 hydrogen bonds, 10 are between CLIP main chain and DR3, and one is between a CLIP side chain (T95) and DR3. The hydrogen bond from  $\alpha 76R$  to the carbonyl of the CLIP residue Q100 is partly obscured. CLIP (yellow) residues P87, T95 and P96 are shown fully, whereas other CLIP residues are displayed only up to C $\beta$ . Only DR3 (silver) residues making hydrogen bonds with CLIP are shown. The view is from above the binding site with the  $\alpha_1$  domain  $\alpha$ -helix at the top and the  $\beta_1$  domain  $\alpha$ -helix (both in red Ca trace) at the bottom. *b*, Six side chains of CLIP fit in pockets formed by DR3. CLIP is shown as bonds whereas DR3 is shown as a surface. View as in *a*. Contacts with DR3 bury 58% of the surface area of CLIP P87–A101. Pockets 1 and 9 bury 190 and 175 Å<sup>2</sup> of CLIP solvent-accessible surface area, respectively; pockets 6 and 7, smaller cavities, bury 122 and 160 Å<sup>2</sup>; pocket 3, buries only 118 Å<sup>2</sup>, and pocket 4 buries 99 Å<sup>2</sup>. The methylene group of arginine 92 of CLIP, which does not fit into a pocket, nevertheless contacts the surface of DR3 and positions the guanidinium group 3.2 Å from  $\beta 81$  His.

**METHODS.** Calculation of solvent-accessible surface areas were performed with a probe radius of 1.4 Å. Residues contacting CLIP in each pocket are: pocket 1,  $\alpha 24F$ ,  $\alpha 31I$ ,  $\alpha 32F$ ,  $\alpha 43W$ ,  $\alpha 52A$ , main chain of  $\alpha 53S$ ,  $\alpha 54F$ ,  $\beta 81H$ ,  $\beta 82N$ ,  $\beta 85V$ ,  $\beta 86V$ ; pocket 3,  $\alpha 9Q$ ,  $\alpha 22F$ ,  $\alpha 24F$ ,  $\alpha 54F$ ,  $\alpha 55E$ ,  $\alpha 58G$ ,  $\alpha 59A$ ,  $\alpha 61A$ ,  $\alpha 62N$ ,  $\beta 78Y$ ; pocket 4,  $\alpha 9Q$ ,  $\alpha 62N$ ,  $\beta 13S$ ,  $\beta 74R$ ,  $\beta 78Y$ ; pocket 6,  $\alpha 9Q$ ,  $\alpha 11E$ ,  $\alpha 62N$ ,  $\alpha 65V$ ,  $\alpha 66D$ ,  $\alpha 69N$ ,  $\beta 11S$ ,  $\beta 13S$ ,  $\beta 28D$ ,  $\beta 30Y$ ,  $\beta 71K$ ,  $\beta 74R$ ; pocket 7,  $\alpha 65V$ ,  $\alpha 69N$ ,  $\beta 30Y$ ,  $\beta 47F$ ,  $\beta 61W$ ,  $\beta 67L$ ,  $\beta 70Q$ ,  $\beta 71K$ ,  $\beta 74R$ ; pocket 9,  $\alpha 69N$ ,  $\alpha 72I$ ,  $\alpha 73M$ ,  $\alpha 76R$ ,  $\beta 9E$ ,  $\beta 30Y$ ,  $\beta 37N$ ,  $\beta 38V$ ,  $\beta 57D$ ,  $\beta 61W$ . R92 of CLIP does not form a pocket contact, but 121 Å<sup>2</sup> of its surface area is rendered inaccessible to solvent by contact with DR3. This figure was made with GRASP<sup>48</sup>.

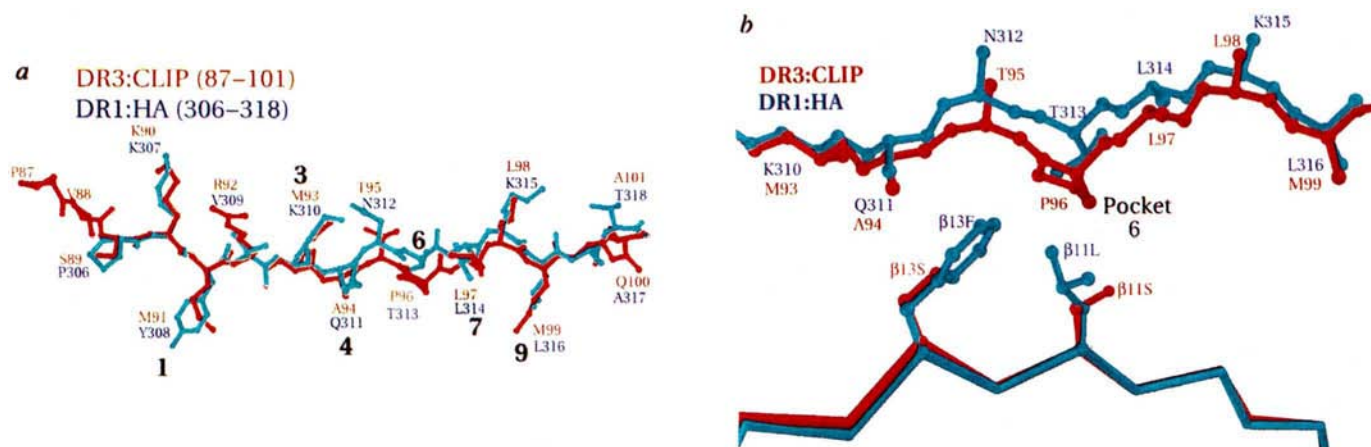


FIG. 3 Conformation of CLIP. *a*, Superposition of CLIP 87-101 with HA 306-318, based on least-squares  $\alpha$ -carbon overlap<sup>49</sup> of the  $\alpha_1\beta_1$  peptide-binding domains of DR3 and DR1. Both peptides have polyproline type II conformations, with mean  $\phi$ ,  $\psi$  angles of  $-87 \pm 20^\circ$ ,  $140 \pm 14^\circ$  for CLIP, and  $-82 \pm 15^\circ$ ,  $139 \pm 16^\circ$  for HA. View as in Fig. 1*b*. The positions of pockets are indicated. CLIP extends two residues further at the N terminus than HA. Although the first 5 overlapping residues (S89-M93 of CLIP, P306-K310 of HA) have similar main-chain paths (r.m.s. deviation of 0.31 Å), the last 8 residues (A94-A101 of CLIP, Q311-T318 of HA) differ (r.m.s. deviation of 1.12 Å), with the largest

displacement at pocket 6. *b*, Superposition of DR3-CLIP (red) and DR1-HA (blue) at pocket 6 showing the displacement of P96 of CLIP and the allelic differences between DR3 and DR1 at  $\beta 11$  and  $\beta 13$ . The peptides are represented as main-chain traces with the position of side chains indicated by their C $\beta$  atoms; main-chain carbonyls are not shown. Side chains for the residues in pocket 6, P96 for CLIP and T313 for HA peptide, are shown. The side chains for  $\beta 11$  and  $\beta 13$  are shown in a ball-and-stick model, and the  $\beta$ -strand containing these residues is shown in a C $\alpha$  trace.

same pockets in the DR3 binding site as those observed in DR1 (ref. 22) (Fig. 2*b*).

### Hydrogen bonds

The conformation of the main chain of bound CLIP is extended but twists such that side chains project from the main chain about every  $128^\circ$ , as in DR1-HA. This allows 17 hydrogen bonds to form between atoms of CLIP and DR3 (Fig. 2*a*), of which 11 are to MHC atoms conserved in class II MHC sequences. The ten hydrogen bonds between the CLIP main-chain atoms and DR3 were observed in the DR1-HA complex and predicted to play a universal role in binding peptides to class II molecules<sup>22</sup>. Overall, the pattern of 17 hydrogen bonds in DR3-

CLIP is essentially identical to the 18 bonds in the DR1-HA complex<sup>22</sup> (compare Fig. 2*a* here and Fig. 3 from ref. 22).

### CLIP side chains and DR3 pockets

A series of pockets that accommodate the side chains of CLIP in the DR3 peptide binding site (labelled 1, 3, 4, 6, 7 and 9 in Fig. 2*b*) are positioned in the same way as those in DR1 (ref. 22). Pockets 1 and 9, deep cavities at either ends of the site, hold CLIP methionines 91 and 99 (Fig. 2*b*); pockets 6 and 7, smaller cavities in the centre of the site, hold CLIP proline 96 and leucine 97; and pocket 3, which is more like a non-polar shelf, holds CLIP methionine 93 and leaves 30% of the methionine's surface exposed. In the DR1-HA complex, that pocket holds the methyl-

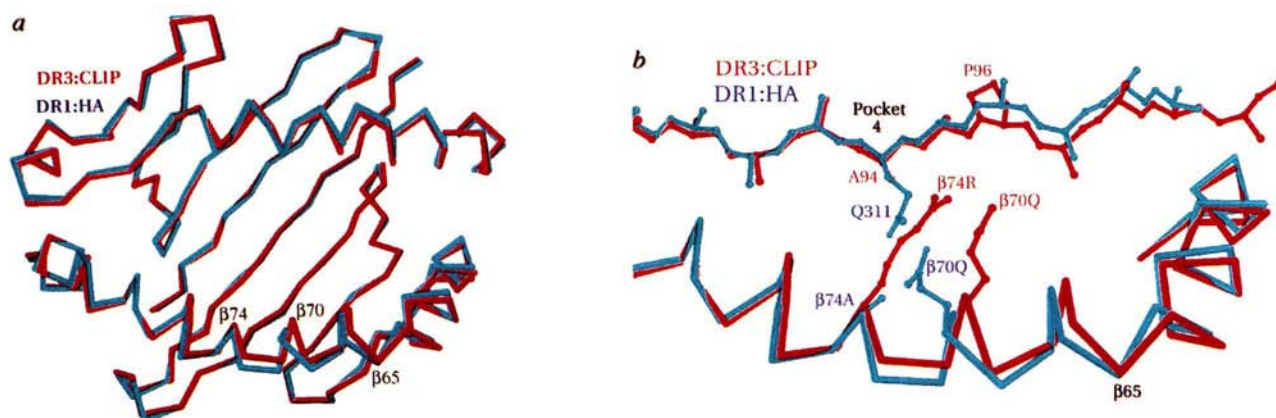


FIG. 4 A comparison of DR3 and DR1 from the CLIP and HA complexes. *a*, Superposition of the  $\alpha_1\beta_1$  peptide-binding domains of DR3-CLIP (red) and DR1-HA (blue) (residues  $\alpha 5$ - $\alpha 80$  and  $\beta 5$ - $\beta 94$ ). View as in Fig. 2. The  $\alpha_1\beta_1$  peptide binding domains are very similar except for an  $\alpha$ -helical region between  $\beta 65$  and  $\beta 74$ . (Not shown are the immunoglobulin-like  $\alpha_2$  and  $\beta_2$  domains which differ between the two complexes by small rotations and translations, probably as a result of different crystal packing forces. Relative to DR1-HA, the  $\alpha_2$  domain of DR3-CLIP is rotated  $3^\circ$  and translated 1 Å, and the  $\beta_2$  domain is rotated  $5^\circ$  and translated 2 Å, as determined by a least-squares superposition based

on the C $\alpha$  positions of the  $\alpha_1\beta_1$  peptide-binding domains. Individually, the  $\alpha_2$  domains of DR3 and DR1 superimpose with an r.m.s. deviation of 0.56 Å and the  $\beta_2$  domains with an r.m.s. deviation of 0.63 Å. Similar shifts in the immunoglobulin-like domains have been observed for class I MHC<sup>50</sup>). *b*, Superposition of DR3-CLIP (red) with DR1-HA (blue) at the  $\beta 65$ - $\beta 74$  region of the  $\beta_1$  domain  $\alpha$ -helix. View as in Fig. 2. The main-chain atoms of the peptides are shown, except for main-chain carbonyls. Side-chain positions are indicated by the C $\beta$  atoms. The side chains of Q311 of HA peptide and for reference, P96 of CLIP, are shown.

TABLE 3 MHC class II pockets involved in binding CLIP

| Conserved pockets | CLIP residue | Area conserved (%) * | DR3 residues: conserved or conservatively substituted in class II MHC †                                                                                                                              |
|-------------------|--------------|----------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1                 | M91          | 75                   | <b><math>\alpha</math>32F, <math>\alpha</math>43W</b> , main chain <b><math>\alpha</math>53S, <math>\alpha</math>54F, <math>\beta</math>81H, <math>\beta</math>82N, <math>\beta</math>85 (V/A/L)</b> |
| 3                 | M93          | 60                   | <b><math>\alpha</math>22 (Y/F), <math>\alpha</math>54F, <math>\alpha</math>55 (D/E), <math>\alpha</math>62N, <math>\beta</math>78 (Y/F/M/I)</b>                                                      |
| 7                 | L97          | 58                   | <b><math>\alpha</math>65 (V/I), <math>\alpha</math>69N, <math>\beta</math>47 (Y/F), <math>\beta</math>61W, <math>\beta</math>67 (L/I/F/V)</b>                                                        |
| 9                 | M99          | 54                   | <b><math>\alpha</math>69N, <math>\alpha</math>73 (M/L/A/V), <math>\alpha</math>76R, <math>\beta</math>38 (V/L/A), <math>\beta</math>61W</b>                                                          |
| Variable pockets  |              |                      |                                                                                                                                                                                                      |
| 6                 | P96          | 41                   | <b><math>\alpha</math>62N, <math>\alpha</math>65 (V/I), <math>\alpha</math>69N</b>                                                                                                                   |
| 4                 | A94          | 23                   | <b><math>\alpha</math>62N, <math>\beta</math>78 (Y/V/M/I)</b>                                                                                                                                        |

\* Percentage of DR3-accessible surface area buried by contact with side-chain atoms of CLIP that is composed of residues conserved or conservatively substituted in class II MHC sequences. A probe radius of 1.4 Å was used to calculate solvent-accessible surface area (M. Handschumacher and F. Richards, personal communication). The surface area of DR3 buried by each residue of CLIP was calculated and partitioned with respect to whether the buried DR3 residue was conserved or conservatively substituted as opposed to non-conserved in class II MHC sequences. For example, of the 87.6 Å<sup>2</sup> of DR3 buried by side-chain atoms of M91 in pocket 1, 65.4 Å<sup>2</sup> or 75% derives from residues conserved or conservatively substituted in class II MHC sequences, while the remaining 22.2 Å<sup>2</sup> derives from residues non-conserved in class II MHC sequences.

† Residues conserved (in bold) or conservatively substituted in class II MHC sequences that are buried in the DR3-CLIP complex by contact with side-chain atoms of CLIP. After each conservatively substituted residue is indicated the range of substitutions found in human class II MHC.

ene groups of a peptide lysine<sup>22</sup>. Pocket 4 of DR3 holds the side-chain methyl group of CLIP Ala 94. This pocket appears to have collapsed slightly (see below) relative to the size it may exhibit when holding an aspartic acid side chain in the common DR3 peptide motif<sup>27,28</sup>. The register of CLIP with respect to the protein pockets is in accordance with that proposed from substitution studies of CLIP<sup>11,13</sup>.

Peptides that bind to DR3 exhibit sequence preferences at position 1 where non-polar residues are preferred (M, L, V, I, F, Y), and position 4 where negatively charged aspartic acid is preferred<sup>27,28</sup>. This binding motif correlates well with the non-polar nature of pocket 1 and the positively charged arginine ( $\beta$ 74) in pocket 4 of DR3. Small rearrangements must be possible in both pockets as model building indicates that tyrosine (Y) could not fit in pocket 1 or aspartic acid (D) in pocket 4 without slight movements of DR3 side chains. Some peptides that bind DR3 exhibit a second motif such that, when the residues in pockets 1 and 4 are not ideal (A in 1; N, Q, E, S, T in 4), peptide position 6 is often positively charged (K, H, R). This appears to correlate with the negative charge in pocket 6 from  $\beta$ 28 aspartic acid of DR3.

### Comparison of DR3-CLIP with DR1-HA

Overall, CLIP appears to fit as closely against DR3 as the HA peptide does against DR1. Buried in the CLIP-DR3 complex are 1,386 Å<sup>2</sup> of solvent-accessible surface area of CLIP and 904 Å<sup>2</sup> of DR3, comparable to the 1400 Å<sup>2</sup> of the HA peptide and 930 Å<sup>2</sup> of the DR1 in DR1-HA<sup>22</sup>.

The path of the peptide main chain is nearly identical for CLIP and HA peptide in the first half of the MHC molecule's binding site, but diverges slightly in the second half at pocket 6 (Fig. 3a), where the main chain of CLIP is 1.8 Å closer to the  $\beta$ -sheet floor of the binding site. The difference of two small

amino acids, serines at  $\beta$ 11 and  $\beta$ 13 in DR3, for two large residues, leucine ( $\beta$ 11) and phenylalanine ( $\beta$ 13) in DR1, allows the peptide in DR3 to get slightly deeper into the site (Fig. 3b). A proline, CLIP P96, in pocket six also contributes as its cyclic side chain does not protrude from the peptide backbone.

There is only one small conformation difference in the DR backbone structures between DR3-CLIP and DR1-HA. Overall the  $\alpha$ <sub>1</sub> $\beta$ <sub>1</sub> peptide-binding domains superimpose very well with an r.m.s. deviation of 0.51 Å in  $\alpha$ -carbon positions (Fig. 4a). The small difference occurs in the  $\beta$ -chain  $\alpha$ -helix between residues  $\beta$ 65 and  $\beta$ 74. The result of differences in the locations of  $\beta$ 70 and  $\beta$ 74 in DR1 and DR3 is that the  $\beta$ -chain  $\alpha$ -helix appears to collapse slightly (1.5 Å) towards the peptide in DR3-CLIP, probably because the peptide side chain in pocket 4 is small (Fig. 4b). Modelling an aspartic acid at peptide position 4 requires that  $\beta$ 74R is moved to open pocket 4 slightly.

### Complexes with other class II MHCs

Many of the same binding interactions observed in the DR3-CLIP complex are probably found in CLIP complexes with other class II MHC molecules. The four large CLIP side chains M91, M93, L97 and M99 are bound in those DR3 pockets which have the greatest proportion of their buried surface areas formed by residues conserved or conservatively substituted in class II molecules (Table 3). Thus these CLIP residues may fit into most class II binding sites. The two smallest CLIP side chains, A94 and P96, are bound in those pockets with linings that are the most variable in chemical character in class II molecules. Pocket 4, which holds A94 of CLIP, is extremely variable (23% conserved). Similarly, 41% of the surface area of pocket 6, which holds P96, is formed by residues that undergo non-conservative substitutions in class II molecules. The occurrence of small side chains at positions 94 and 96 in CLIP (A94, P96) may minimize unfavourable contacts in pockets 4 and 6 of most other class II molecules.

### Implications for II activities

After proteolytic degradation of Ii *in vivo*, CLIP is seen, from the structure of DR3-CLIP, to bind in the  $\alpha\beta$  peptide binding site using the same network of hydrogen bonds and contacts to protein pockets as an antigenic peptide. This provides a mechanism for the chaperoning role of Ii, because CLIP in the binding site would be expected to stabilize  $\alpha\beta$  against aggregation *in vivo*<sup>9</sup>, as antigenic peptides do *in vitro*<sup>29,30</sup>. Genetic truncation experiments<sup>8-10</sup>, pulse-chase experiments<sup>6</sup>, antibody binding<sup>31</sup> (E.M. and Y. Paterson, manuscript in preparation) and T-cell<sup>31</sup> recognition experiments have already provided evidence that Ii binds to MHC class II molecules by inserting CLIP into the peptide binding site. The region surrounding CLIP in a soluble, recombinant Ii has recently been shown to be unstructured, as determined by nuclear magnetic resonance and proteolytic sensitivity<sup>32,33</sup>, making CLIP available for binding as observed here. The presence of CLIP in the peptide binding site also explains how Ii blocks peptide binding *in vitro*<sup>9,10</sup>. For those class II molecules (for example, I-A<sup>k</sup> and DR52a) which have a low affinity for CLIP<sup>12</sup>, the C-terminal domain of Ii may augment the affinity<sup>33</sup>.

Most studies of CLIP binding to DR<sup>11-13,34</sup> have suggested the binding mode observed here, and the effects on CLIP binding by mutations at  $\alpha$ 58 and  $\beta$ 86 in a mouse class II MHC<sup>14</sup> are also consistent with the observation that CLIP binds near both these residues (Fig. 2b legend). A suggestion that the C-terminal part of CLIP makes contacts to DR2 other than those made by conventional groove-binding peptides or as described for CLIP<sup>11-13,34</sup> could have resulted from the use of a short peptide lacking the P1 anchor and/or the addition of an N-terminal fluorescent probe to that peptide<sup>15</sup>.

The details of the atomic contacts between CLIP and DR3 are consistent with the *in vivo* stability inferred from the ability of DR3-CLIP to traverse the endosome and survive on the cell



surface in DM-deficient and wild-type cells<sup>2-6,35,36</sup>. The half-life of soluble DR3-CLIP isolated from DM-deficient cells is almost 4 hours in conditions (pH 4.5 and 37 °C) mimicking the endosomal environment (Table 2). This is longer than the 1–2 hours required for class II MHC maturation<sup>37</sup>, and suggests that the structure of DR3-CLIP represents the structure of the complex in the endosome, before DM-mediated exchange.

CLIP's nine N-terminal residues (L81–S89) have been suggested to increase CLIP dissociation from soluble DR1 (ref. 26) and from detergent-solubilized DR2 (ref. 15). In the latter case, the fast off-rate reported was probably the result of using a detergent (Zwittergent), which has a  $\beta$ -OG-like destabilizing effect<sup>6</sup> on DR3-CLIP (data not shown). In the former case, experiments are needed to determine whether it is the ordered or disordered residues at the terminus that confer this reported property.

### Implications for DM exchange reaction

The finding that CLIP binds to DR3 like HA binds DR1 has implications for DM's exchange mechanism<sup>18-20</sup>. Although previous observations of a difference in SDS stability and antibody binding<sup>2-5,21</sup> had been interpreted to indicate an altered DR3-CLIP conformation, the structure shows only a small difference between DR3-CLIP and DR1-HA: a segment of the  $\beta$ -chain  $\alpha$ -helix ( $\beta$ 65–74) in DR3-CLIP is displaced about 1.5 Å towards the peptide binding site (Fig. 4), forming a slightly more 'closed' site rather than the more 'open' site that SDS instability was thought to indicate. This 'closing' may result from movement of

$\beta$ 74 Arg and the collapse of pocket 4 to fit around alanine 94 of CLIP, a peptide position normally occupied by aspartic acid in peptides that bind DR3. Replacement of alanine 94 by aspartic acid results in SDS-stable complexes<sup>11</sup>. The functional consequences, if any, of the lack of SDS stability (and of instability in  $\beta$ -OG<sup>6</sup>) of DR3-CLIP are not clear. Some antigenic peptides, even ones that are recognized by T cells, form complexes that are SDS unstable<sup>30,38</sup> and, importantly, some DR-CLIP complexes, such as DR1-CLIP, are stable in SDS<sup>35</sup> but readily exchanged by DM<sup>18,20</sup>. This segment of DR3's  $\beta$  chain probably also accounts for the difference in antibody recognition between DR3-CLIP and other DR3-peptide complexes, as the binding site of an antibody, NDS-9, that recognizes DR3-peptide complexes but not DR3-CLIP<sup>4</sup> has been mapped to  $\beta$ 73, 74 and 77 (ref. 39) (Fig. 4).

The observation that DR3-CLIP and DR1-HA are highly similar in structure suggests that DM does not discriminate between CLIP and HA by binding to an alternate equilibrium conformer of DR that would be present in DR-CLIP but not in DR-HA. Instead, the increase in the rate of peptide dissociation by DM suggests that DM may recognize and stabilize a strained, transition-state-like conformer of DR, lowering the free-energy barrier for CLIP dissociation. This conformer may exist for both DR-HA and DR-CLIP, but stabilization by DM may result in the preferential dissociation of peptides lacking optimal anchor residues, such as CLIP<sup>18,20</sup>. The structure of DR3-CLIP, a substrate for DM-mediated exchange, suggests that DM functions on a transient state of class II MHC rather than on a new equilibrium conformation. □

Received 15 May; accepted 24 October 1995.

- Cresswell, P. A. *Rev. Immun.* **12**, 259–293 (1994).
- Riberdy, J. M., Newcomb, J. R., Surman, M. J., Barbosa, J. A. & Cresswell, P. *Nature* **360**, 474–477 (1992).
- Sette, A. et al. *Science* **258**, 1801–1804 (1992).
- Mellins, E. et al. *J. exp. Med.* **179**, 541–549 (1994).
- Monji, T., McCormack, A. L., Yates, J. R. & Pious, D. J. *exp. Med.* **153**, 4468–4477 (1994).
- Avva, R. & Cresswell, P. *Immunity* **1**, 763–774 (1994).
- Xu, M., Capraro, G. A., Daibata, M., Reyes, V. E. & Humphreys, R. E. *Molec. Immun.* **31**, 723–731 (1994).
- Freisewinkel, I. M., Schenck, K. & Koch, N. *Proc. natn. Acad. Sci. U.S.A.* **90**, 9703–9706 (1993).
- Romagnoli, P. & Germain, R. N. *J. exp. Med.* **180**, 1107–1113 (1994).
- Bijlmakers, M. E., Benaroch, P. & Ploegh, H. L. *J. exp. Med.* **180**, 623–629 (1994).
- Malcherek, G., Gnau, V., Jung, G., Rammensee, H.-G. & Melms, A. *J. exp. Med.* **181**, 527–536 (1995).
- Sette, A., Southwood, S., Miller, J. & Apella, E. *J. exp. Med.* **181**, 677–683 (1995).
- Geluk, A., van Meijgaarden, K. E., Drijfhout, J. W. & Ottenhof, T. H. M. *Molec. Immun.* **32**, 975–981 (1995).
- Gautam, A. N., Pearson, C., Quinn, V., McDevitt, H. O. & Milburn, P. J. *Proc. natn. Acad. Sci. U.S.A.* **92**, 335–339 (1995).
- Kropshofer, H., Vogt, A. B. & Hämmerling, G. J. *Proc. natn. Acad. Sci. U.S.A.* **92**, 8313–8317 (1995).
- Morris, P. et al. *Nature* **368**, 551–554 (1994).
- Fling, S. P., Arp, B. & Pious, D. *Nature* **368**, 554–558 (1994).
- Sloan, V. S. et al. *Nature* **375**, 802–806 (1995).
- Denzin, L. K. & Cresswell, P. *Cell* **82**, 155–165 (1995).
- Sherman, M. A., Weber, D. A. & Jensen, P. E. *Immunity* **3**, 197–205 (1995).
- Mellins, E. et al. *Nature* **343**, 71–74 (1990).
- Stern, L. J. et al. *Nature* **368**, 215–221 (1994).
- Brown, J. H. et al. *Nature* **368**, 33–39 (1993).
- Jardetzky, T. S. et al. *Nature* **368**, 711–718 (1994).
- Kim, J., Urban, R. G., Strominger, J. L. & Wiley, D. C. *Science* **266**, 1870–1874 (1994).
- Urban, R. G., Chicz, R. M. & Strominger, J. L. *J. exp. Med.* **180**, 751–755 (1994).
- Malcherek, G. et al. *Int. Immun.* **5**, 1229–1237 (1993).
- Geluk, A. et al. *J. Immun.* **152**, 5742–5748 (1994).

- Stern, L. J. & Wiley, D. C. *Cell* **68**, 465–477 (1992).
- Germain, R. N. & Rinker, A. G. *Nature* **363**, 725–728 (1993).
- Morkowski, S. et al. *J. exp. Med.* **182**, 1403–1413 (1995).
- Jasanoff, A., Park, S.-J. & Wiley, D. C. *Proc. natn. Acad. Sci. U.S.A.* **92**, 9900–9904 (1995).
- Park, S.-J., Sadegh-Nasseri, S. & Wiley, D. C. *Proc. natn. Acad. Sci. U.S.A.* **92**, 11289–11293 (1995).
- Lee, C. & McConnell, H. *Proc. natn. Acad. Sci. U.S.A.* **92**, 8269–8273 (1995).
- Chicz, R. M. et al. *Nature* **358**, 764–768 (1992).
- Chicz, R. M. et al. *J. exp. Med.* **178**, 27–47 (1993).
- Tulp, A., Verwoerd, D., Dobberstein, B., Ploegh, H. L. & Pieters, J. *Nature* **369**, 120–126 (1994).
- Nelson, C. A., Petzold, S. J. & Unanue, E. R. *Nature* **371**, 250–252 (1994).
- Maurer, D. & Gorski, J. *J. Immun.* **146**, 621–626 (1991).
- Gorga, J. C., Horejsi, V., Johnson, D. R., Raghupathy, R. & Strominger, J. L. *J. biol. Chem.* **262**, 16087–16094 (1987).
- Gorga, J. C., Brown, J. H., Jardetzky, T., Wiley, D. C. & Strominger, J. L. *Res. Immun.* **142**, 401–407 (1991).
- Collaborative Computational Project, N. *Acta crystallogr.* **D50**, 760–763 (1994).
- Navaza, J. *Acta crystallogr.* **A50**, 157–163 (1994).
- Brünger, A. T. *Nature* **355**, 472–475 (1992).
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47**, 110–119 (1991).
- Luzzati, V. *Acta crystallogr.* **5**, 802–810 (1952).
- Huang, C. C., Petterson, E. F., Klein, T. E., Ferrin, T. E. & Langridge, R. J. *molec. Graphics* **9**, 230–236 (1991).
- Nicholls, A., Sharp, K. A. & Honig, B. *Proteins* **11**, 281–296 (1991).
- Kabsch, W. *Acta crystallogr.* **A34**, 827–828 (1978).
- Madden, D. R., Gorga, J. C., Strominger, J. L. & Wiley, D. C. *Cell* **70**, 1035–1048 (1992).

ACKNOWLEDGEMENTS. We thank P. Rosenthal, E. Collins, L. Stern, G. Weiss, A. Taurog, D. Garboczi, S. Ray, S.-J. Park, J. Kim and the staff of the Cornell High Energy Synchrotron Source MacCHESS for help in data collection; R. Chicz for mass spectrometry; P. Pietras for large-scale tissue-culture growth; E. Petterson for help with MidasPlus; and C. Mueller, F. Rey, A. Jasanoff, S.-J. Park and others for discussion. P.G. was supported by the Irvington Institute for Medical Research; M.A. and E.M. by NIH; and D.C.W. is an investigator with the Howard Hughes Medical Institute. Coordinates will be deposited in the Protein Data Bank (Brookhaven National Laboratory, Upton, New York), and are available before their release by e-mail (ghosh@xtal220.harvard.edu).



## Discovery of a cool brown dwarf

T. Nakajima\*, B. R. Oppenheimer\*, S. R. Kulkarni\*,  
D. A. Golimowski†, K. Matthews\* & S. T. Durrance†

\* Palomar Observatory 105-24, California Institute of Technology,  
Pasadena, California 91125, USA

† Department of Physics and Astronomy,  
The Johns Hopkins University, Baltimore, Maryland 21218, USA

**BROWN dwarfs are star-like objects with masses less than 0.08 times that of the Sun, which are unable to sustain hydrogen fusion in their interiors<sup>1-4</sup>. They are very hard to detect, as most of the energy of gravitational contraction is radiated away within  $\sim 10^6$  yr, leaving only a very low residual luminosity. Accordingly, almost all searches for brown dwarfs have been directed towards clusters of young stars—a strategy that has recently proved successful<sup>5,6</sup>. But there are only modest observable differences between young brown dwarfs and very low-mass stars, making it difficult to identify the former without appealing to sophisticated models<sup>7</sup>. Older brown dwarfs should have a more distinctive appearance, and if they are companions to nearby stars, their luminosity can be determined unambiguously. Here we report the discovery of a probable companion to the nearby star Gl229, with no more than one-tenth the luminosity of the least luminous hydrogen-burning star. We conclude that the companion, Gl229B, is a brown dwarf with a temperature of less than 1,200 K, and a mass  $\sim 20$ –50 times that of Jupiter.**

Three years ago, we initiated a search<sup>8</sup> for brown-dwarf companions to stars less than 15 pc away. With this strategy, we utilize the precisely known distances to these stars to determine accurately the total luminosity of any discovered companions. In all models of low-mass stars<sup>4</sup>, the lowest luminosity of any hydrogen-burning star is  $L_{\min} \geq 10^{-4} L_{\odot}$ . Brown dwarfs of age  $10^9$  yr have luminosities well below  $L_{\min}$  and are therefore clearly distinguishable<sup>3</sup> from low-mass stars. Consequently, we have begun the 'Palomar-Byr' survey, a survey of all nearby stars with ages of about one billion ( $10^9$ ) years. These stars have either small space motion<sup>9</sup> or have active coronae (as inferred from

copious H $\alpha$ <sup>10</sup> or X-ray emission<sup>11</sup>), both of which are indicators of stellar youth. The survey consists of both optical coronagraphic imaging and infrared direct imaging of the stars.

During the past year, we have observed about 100 stars of this sample. Preliminary analysis of the data revealed a very intriguing proper-motion companion to the nearby star, Gl229 (also known as HD42581 and LHS1827; spectral type MIV;  $M_V = 9.3$ , parallax<sup>9</sup>  $0.1749 \pm 0.0045$  arcsec or a distance of 5.7 pc; see ref. 12 for further details). This star was selected from a list of stars with known low space motion<sup>9</sup>. Based on its kinematics it has been classified as a 'young' disk star, age  $\sim 10^9$  yr. However, the star exhibits H $\alpha$  absorption with an equivalent width of 0.5 Å (G. Basri, J. Gizis and N. Reid, personal communication). All of the MI stars in the Pleiades (age  $\sim 10^8$  yr) show H $\alpha$  in emission. However, some stars with  $M_V > 9.5$  in the Hyades<sup>13</sup> (age  $\sim 8 \times 10^8$  yr) show H $\alpha$  absorption even as strongly as Gl229. This suggests that Gl229 is possibly at least as old as the Hyades. The kinematics make it unlikely that the object is older than the Sun.

We observed Gl229 with the Johns Hopkins University's Adaptive Optics Coronagraph<sup>14</sup> at the Cassegrain focus of the Palomar 60-inch telescope on the nights of 27 and 29 October 1994 UT. A compact source, hereafter Gl229B, was detected  $7.78 \pm 0.1$  arcsec from Gl229 at a southeasterly position angle  $PA = 163^\circ \pm 1^\circ$ . Here PA is measured from north to east (Fig. 1). Exposures in three different bandpasses, Thuan-Gunn  $r$ ,  $i$  and  $z$ , showed that the object was very red (Fig. 2). Infrared data of Gl229B in the J, H, K<sub>s</sub> and K bands were obtained at the Hale 200-inch telescope on the nights of 14 and 15 September 1995 UT. During October 3–8 1995 UT, Gl229 was observed again with the Adaptive Optics Coronagraph (AOC) at the Palomar 60-inch telescope. During this run, we also observed the field centred on the star  $\theta_1$  Orionis. This field has a number of stars with well determined positions<sup>15</sup> with respect to  $\theta_1$  Ori. These observations enabled us to derive the pixel scale (0.118 arcsec per pixel) and also determine the orientation of the CCD (charge-coupled device) detector attached to the AOC. The orientation was as expected: the CCD columns were oriented north-south to within  $0.1^\circ$ .

FIG. 1 Optical and infrared images of Gl229. Each field of view is  $25'' \times 25''$ , with north up and east to the left. Top left,  $r$ -band image; top right,  $i$ -band image, bottom left,  $z$ -band image; bottom right, K<sub>s</sub>-band image. The logarithm of the data is displayed in all four panels. The brown-dwarf companion Gl229B is located 7.6 arcsec to the SSE of Gl229A. The  $r$ ,  $i$  and  $z$  images were obtained on epoch 1994.82 with the Adaptive Optics Coronagraph (AOC) and the Palomar 60-inch telescope. The AOC<sup>14</sup> contains an image stabilizer for improved resolution and a pupil apodizer for suppression of light diffracted by the telescope aperture. The K<sub>s</sub> image was obtained on epoch 1995.70 at the Hale 200-inch telescope. The infrared camera, located at the  $f/70$  Cassegrain focus, contains reimaging optics and a Santa Barbara Research Corporation InSb detector. The secondary mirror used signals from a nearby field star to reduce image wander. A cryogenically cooled occulting disk masked Gl229A. The triangular support for this disk is seen at the left of the K<sub>s</sub> image.

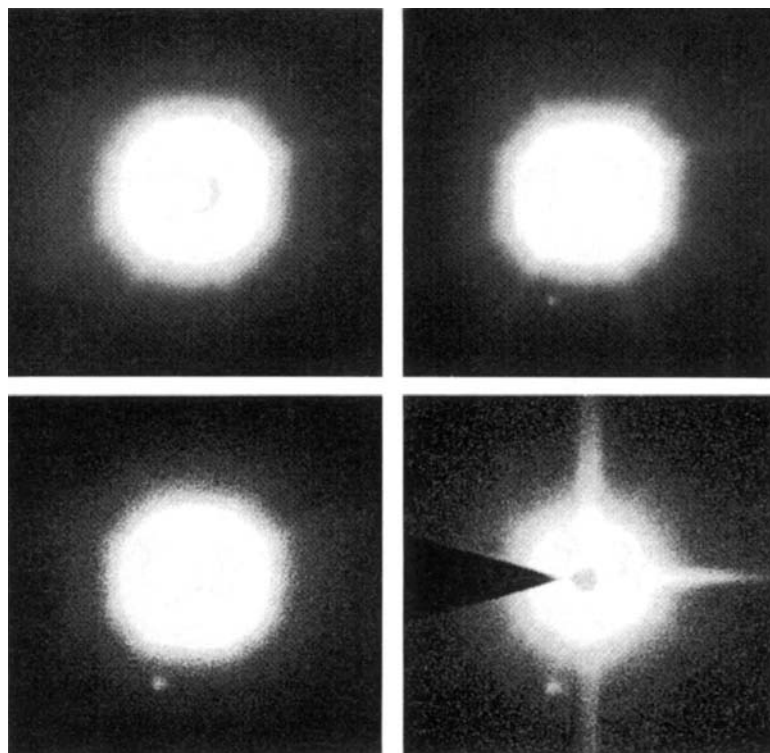
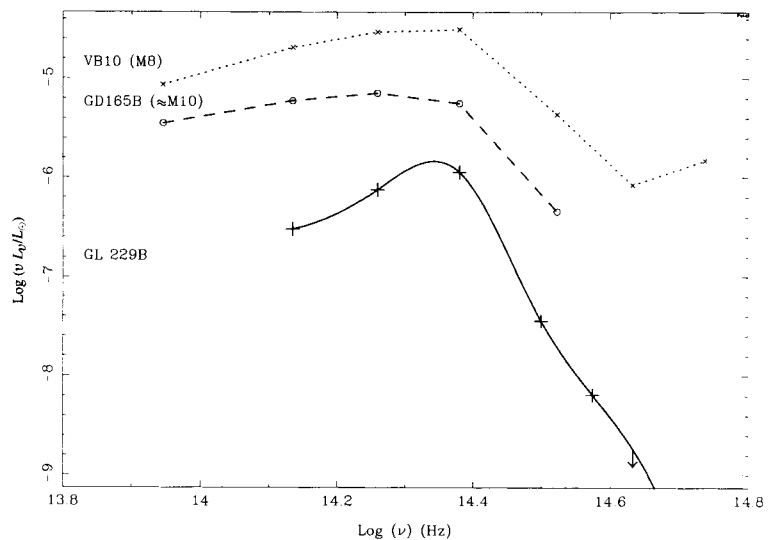


FIG. 2 Broad-band spectra of Gl229B, GD165B and VB10. The ordinate is the logarithm of  $\nu L_\nu$  in units of solar luminosity where  $\nu$  is the frequency and  $L_\nu$  the luminosity per unit frequency. The spectral band in which most of the energy emerges is best indicated by this quantity. The observed magnitudes of Gl229B are  $r > 22.1$  (a  $10\sigma$  upper limit),  $i = 20.0 \pm 0.2$ ,  $z = 17.8 \pm 0.3$ ,  $J = 13.5 \pm 0.1$ ,  $H = 13.2 \pm 0.1$ ,  $K = 13.4 \pm 0.1$  and  $K_s = 13.3 \pm 0.1$ . The absolute magnitudes are;  $M_i > 23.3$ ,  $M_i = 21.2 \pm 0.2$ ,  $M_z = 19.0 \pm 0.3$ ,  $M_J = 14.7 \pm 0.1$ ,  $M_H = 14.4 \pm 0.1$ ,  $M_{K_s} = 14.5 \pm 0.1$ ,  $M_K = 14.6 \pm 0.1$ , with a distance modulus of  $(M - m) = +1.21 \pm 0.05$ . The solid line is a cubic spline interpolation of the observed points with an extension to infinite frequency. The data for GD165B and the late M dwarf, VB10, are obtained from ref. 22. Gl229B is remarkable for both its low luminosity and its peaked spectral energy distribution.



The infrared image (Fig. 1) showed an object at the position of Gl229B. The  $r$ -band AOC images contain six stars, excluding Gl229, spread over the field of view of  $1 \text{ arcmin}^2$ . We compared the relative positions of these stars in the 1994 and 1995 images. We found that there was very little change in the pixel scale, but that the 1994 image was rotated  $1.8^\circ \pm 0.2^\circ$  with respect to the 1995 image. This comparison also yielded an empirical measure of the root-mean-square error,  $\sigma \sim 1$  pixel in each axis, in the fiducial centre of the star (centroid). The occulting disk had enough transmission that we could satisfactorily centre on Gl229. From 1994 to 1995, we found that Gl229 had moved, with respect to the six field stars, by  $0.3 \pm 1$  pixel to the west and  $4.5 \pm 1$  pixel to the south. This is consistent with the known proper motion<sup>12</sup> of Gl229,  $0.737 \text{ arcsec yr}^{-1}$  and  $\text{PA} = 190.2^\circ$ . Gl229B is not detected in the  $r$ -band AOC image but is easily detected in the  $z$ -band image. Unfortunately, only two of the six field stars are seen in this image. With respect to these, Gl229B had moved  $0 \pm 1.5$  pixel to the west and  $3.7 \pm 1.5$  pixel to the south. Thus within the errors, it appears that Gl229B is a common proper-motion companion of Gl229. It is appropriate to refer to Gl229 as Gl229A.

Our argument of companionship is based on two measurements, obtained about one year apart. It is certain that Gl229B is not a distant object because it has a measurable proper motion. It is conceivable, though improbable, that Gl229B is a foreground object, unrelated to Gl229A. Fortunately, this issue can be resolved through observations. Specifically, measurement of the parallax (easily possible with the Hubble Space Telescope) of Gl229B is the most definitive way to confirm companionship.

To undertake further observations (for example, long-slit spectroscopy) it is necessary to obtain the precise offset of Gl229B with respect to Gl229A. This is best done using the infrared data rather than the AOC data for the following reason. At optical wavelengths, the contrast ( $C$ ) defined as the ratio of the brightness of Gl229A to Gl229B is extremely large (wavelength  $\lambda \sim 0.65 \mu\text{m}$ ,  $C > 10^6$ ;  $\lambda \sim 0.80 \mu\text{m}$ ,  $C = 4 \times 10^5$ ;  $\lambda \sim 0.9 \mu\text{m}$ ,  $C \approx 8 \times 10^4$ ). For this reason, we have to use a coronagraph in order to image Gl229B and Gl229A simultaneously. However, spatial variations in the transparency across the occulting disk will bias the determination of the centroid of Gl229A but not that of Gl229B. An additional complication is the point-spread function varies across the field of view of the coronagraph. The situation is far more favourable at infrared wavelengths where  $C \approx 60$  and thus direct imaging techniques can be used. From a number of short and long exposures we find that Gl229B is located  $7.69 \pm 0.1 \text{ arcsec}$  from Gl229A at  $\text{PA} = 162^\circ \pm 1^\circ$ . The

pixel scale and the orientation of the infrared camera are extremely well determined and known to be very stable.

The magnitudes of Gl229B are given in Fig. 2. The total (or bolometric) luminosity  $L$  can be obtained by integrating the broad-band spectrum shown in Fig. 2. This integration yields  $L \approx 2 \times 10^{-6} L_\odot$ . It is possible that there is considerable emission at frequencies smaller than the lowest measured frequency at K band. However, this method overestimates the total emission due to the presence of opacity between the observed bands. For these reasons, an accurate estimation of the total emission requires guidance from models of brown-dwarf atmospheres. These models are still in their infancy, and they critically rely on effective temperature,  $T_{\text{eff}} = (L/4\pi R^2 \sigma)^{1/4}$ , where  $R$  is the radius and  $\sigma$  is the Stefan-Boltzmann constant. For brown dwarfs<sup>3,4</sup>,  $R \approx 0.1 R_\odot$ . We obtain  $T_{\text{eff}} = 700 \text{ K}$  for Gl229B. Only the models of Tsuji *et al.*<sup>16</sup> are applicable to such a low  $T_{\text{eff}}$ . They predict  $K - L' = 2$  for the observed  $J - K = 0.15$  and  $T_{\text{eff}} \approx 1,000 \text{ K}$ . Incorporating the predicted  $L'$  and extrapolating the low-frequency tail with a Rayleigh-Jeans formula, we obtain  $L \approx 4 \times 10^{-6} L_\odot$ .

The definition of a brown dwarf depends on its mass being less than 80 times the mass of Jupiter. However, with the data at hand, it is not possible to determine the mass of Gl229B from dynamical considerations<sup>17</sup>. But according to models of low-mass stars<sup>4</sup>, the luminosity of the lowest-mass star is  $L_{\text{min}} \approx 10^{-4} L_\odot$ . This is about equal to the luminosity<sup>18,19</sup> of GD165B. Unfortunately, the mass of this object is unknown. It is certainly on the threshold of the star-substellar boundary. The red dwarf with the lowest determined dynamical mass<sup>17</sup> is GJ1245C,  $M = 0.074 \pm 0.013 M_\odot$ . The luminosity of this star is  $\sim 5 \times 10^{-4} L_\odot$ . For Gl229B, we make the conservative assumption of  $L < 10^{-5} L_\odot$  which is more than order of magnitude smaller than the luminosity of either of these two objects. Thus we conclude that Gl229B is certainly a brown dwarf. Below we argue that the peculiar broad-band spectrum of Gl229B constitutes an independent line of reasoning that supports this conclusion.

We now use brown-dwarf cooling models to estimate the mass of Gl229B. According to these models<sup>4</sup>, the luminosity of a brown dwarf older than  $10^8 \text{ yr}$  is

$$L = 3.4 \times 10^{-6} L_\odot t_9^{-1.3} (M/20 M_J)^{2.64}$$

where  $t_9$  is the age of the brown dwarf in billions of years,  $M$  is the mass of the brown dwarf, and  $M_J$  is the mass of Jupiter. Thus we infer a mass of  $\sim 20 \sqrt{t_9} M_J$ . Earlier we argued that  $t_9$  is of the order of unity. Even if  $t_9$  is as high as 6, the estimated mass is still below the maximum mass for brown

dwarfs. All this discussion assumes that G1229A and G1229B are coeval.

The broad-band spectrum of G1229B is unlike that of any known star or for that matter any known celestial object. Specifically, the infrared colour is blue whereas the optical colour is extremely red (Fig. 2). The spectral energy distribution peaks at J band,  $\lambda \approx 1.25 \mu\text{m}$ . This is indeed predicted by recent models of brown-dwarf atmospheres<sup>16,20</sup>. This phenomenon is reproduced by all models regardless of the details of the input molecular opacities. At high pressures and low temperatures, characteristic of atmospheres of brown dwarfs,  $\text{H}_2\text{O}$  and  $\text{CH}_4$  are formed abundantly and these molecules essentially limit the transmission to the near-infrared band of  $1.2 \mu\text{m}$ . The cool temperature results in very little emission below  $1 \mu\text{m}$ . These two effects conspire to produce a peak at J band. Thus we argue that the peculiar broad-band spectrum provides additional evidence

that G1229B is a cool brown dwarf. Elsewhere<sup>21</sup> we report the detection of strong methane absorption in the spectrum of this object which provides the definitive proof that G1229B is a cool object,  $T_{\text{eff}} < 1,200 \text{ K}$ .

We end with some suggestions and speculations. First, searches for isolated brown dwarfs should take advantage of the peculiar colours such as those we see in G1229B. Specifically, the searches should not use infrared colours alone as a discriminant, but rather compare optical flux to any infrared flux. Second, the projected separation of G1229B is 44 astronomical units, which is about the distance from the Sun to Pluto. Thus the orbital motion of G1229B should be  $0.1 \text{ arcsec yr}^{-1}$ . Observations with the Hubble Space Telescope could easily measure this. Newly commissioned infrared interferometers have sufficient precision that the mass of G1229B could be determined over the next few years, provided that the orbit is circular.  $\square$

Received 25 September; accepted 13 November 1995.

1. Kumar, S. S. *Astrophys. J.* **137**, 1121–1126 (1963).
2. Tartar, J. C. in *Astrophysics of Brown Dwarfs* (eds Kafatos, M. C., Harrington, R. S. & Maran, S. P.) 121–138 (Cambridge Univ. Press, 1986).
3. Stevenson, D. J. A. *Rev. Astr. Astrophys.* **29**, 163–193 (1991).
4. Burrows, A. & Liebert, J. *Rev. mod. Phys.* **65**, 301–336 (1993).
5. Basri, G., Marcy, G. W. & Graham, J. R. *Astrophys. J.* (in the press).
6. Rebolo, R., Zapatero Osorio, M. R. & Martin, E. L. *Nature* **377**, 129–131 (1995).
7. Magazzù, A., Martin, E. L. & Rebolo, R. *Astrophys. J.* **404**, L17–L20 (1993).
8. Nakajima, T., Durrance, S. T., Golimowski, D. A. & Kulkarni, S. R. *Astrophys. J.* **428**, 797–804 (1994).
9. Leggett, S. K. *Astrophys. J. Suppl. Ser.* **82**, 351–394 (1992).
10. Hawley, S. L. in *The Bottom of the Main Sequence—and Beyond* (ed. Tinney, C. G.) 224–227 (Springer, Berlin, 1995).
11. Pallaivincini, R. *Astr. Astrophys. Rev.* **1**, 177–207 (1989).
12. Giliese, W. & Jahreiss, H. *Preliminary Version of the Third Catalog of Nearby Stars* (1991).
13. Reid, I. N., Hawley, S. L. & Mateo, M. *Mon. Not. R. astr. Soc.* **272**, 828–842 (1995).
14. Golimowski, D. A., Clampin, M., Durrance, S. T. & Barkhouser, R. H. *Appl. Opt.* **31**, 4405–4416 (1992).

15. Hirschfeld, A. & Sinnott, R. W. *Sky Catalog 2000.0 Vol. 2* (Cambridge Univ. Press, Sky Publishing, Cambridge, 1995).
16. Tsuji, T., Ohnaka, K. & Aoki, W. in *The Bottom of the Main Sequence—and Beyond* (ed. Tinney, C. G.) 45–49 (Springer, Berlin, 1995).
17. Henry, T. J. & McCarthy, D. W. *Astr. J.* **106**, 773–789 (1993).
18. Zuckerman, B. & Becklin, E. E. *Astrophys. J.* **386**, 260–264 (1992).
19. Tinney, C. G., Mould, J. R. & Reid, I. N. *Astr. J.* **105**, 1045–1059 (1993).
20. Allard, F. & Hauschildt, H. P. in *The Bottom of the Main Sequence—and Beyond* (ed. Tinney, C. G.) 32–40 (Springer, Berlin, 1995).
21. Oppenheimer, B. R., Kulkarni, S. R., Matthews, K. & Nakajima, T. *Science* (in the press).
22. Jones, H. R. A., Longmore, A. J., Jameson, R. F. & Mountain, C. M. *Mon. Not. R. astr. Soc.* **267**, 413–423 (1994).

ACKNOWLEDGEMENTS. We thank P. Goldreich, E. S. Phinney, I. N. Reid and M. van Kerkwijk for discussions. Some of the observations were conducted at the Palomar 60-inch telescope, which is jointly owned by the California Institute of Technology and the Carnegie Institution of Washington. At Caltech, this research was initiated from a seed grant from the Flintridge Foundation. S.R.K. is deeply indebted to the Foundation. The construction of the AOC was funded by the Seaver Institute. S.T.D. is grateful for their generous support. Our current activities are supported by the Packard Foundation, the US NSF and NASA. Infrared astronomy at Palomar is supported by the US NSF. B.R.O. is supported by an NSF graduate Fellowship. D.A.G. and S.T.D. thank the Center for Astrophysical Sciences at JHU for its support of this work.

## Taming spatiotemporal chaos with disorder

Y. Braiman<sup>\*†</sup>, John F. Lindner<sup>\*‡</sup>  
& William L. Ditto<sup>\*</sup>

<sup>\*</sup> Applied Chaos Laboratory, School of Physics,  
Georgia Institute of Technology, Atlanta, Georgia 30332, USA

<sup>†</sup> The College of Wooster, Wooster, Ohio 44691, USA

DISORDER and noise in physical systems usually tend to destroy spatial and temporal regularity, but recent research into nonlinear systems provides intriguing counter-examples. In the phenomenon of stochastic resonance<sup>1</sup>, for example, the presence of noise improves the ability of some nonlinear systems to transfer information reliably. Noise can also remove chaos in a model oscillator<sup>2</sup>, and facilitate synchronization in an extended array of bistable elements<sup>3</sup>. Here we explore the use of disorder as a means to control spatiotemporal chaos<sup>4–8</sup> in coupled arrays of forced, damped, nonlinear oscillators. Chaotic behaviour in spatially extended systems, especially in biology and physiology<sup>9,10</sup>, might be amenable to control, as occurs in low-dimensional temporally chaotic systems<sup>11,12</sup>. In our numerical experiments, one- and two-dimensional arrays of identical oscillators behave chaotically, but the introduction of slight, uncorrelated differences between the oscillators induces ordered motion characterized by complex but regular spatiotemporal patterns.

If our oscillators are identical, then our arrays are chaotic. (Nearby trajectories in the phase space of the array diverge exponentially, as measured by a positive Lyapunov exponent). Mak-

ing the oscillators non-identical by introducing disorder into a parameter that characterizes their behaviour, removes the chaos and freezes-in complex spatiotemporal patterns by frequency-locking the motion to harmonics and subharmonics of the forcing. (Frequency-locking is distinct from phase-locking: the former tames chaos whereas the latter need not, as a synchronized array may be chaotic.) This phenomenon appears robust with respect to both the parameter range and the size and realization of the disorder. In fact, when we numerically calculate the leading Lyapunov exponent of the array as a function of the size of the disorder, we find large regions in parameter space where disorder leads to complete frequency-synchronization of the array. Moreover, we have observed this effect in multiple systems.

To demonstrate the phenomenon clearly, we first suggest an intentionally simple example. We consider a one-dimensional array (or chain) of forced, damped, nonlinear pendula governed by the equation

$$ml_n^2 \ddot{\theta}_n + \gamma \dot{\theta}_n = -mgl_n \sin \theta_n + \tau' + \tau \sin \omega t + \kappa(\theta_{n+1} - \theta_n) + \kappa(\theta_{n-1} - \theta_n)$$

where  $n = 1, 2, \dots, N$ , and where the ends are free ( $\theta_0 = \theta_1$  and  $\theta_N = \theta_{N+1}$ ). For simplicity, we take the acceleration due to gravity  $g = 1.0$  and the mass of the pendulum bob  $m = 1.0$ . Other parameters are damping  $\gamma = 0.75$ , length  $l_n = 1.0$ , d.c. torque  $\tau' = 0.7155$ , a.c. torque  $\tau = 0.4$ , angular frequency  $\omega = 0.25$ , coupling  $\kappa = 0.5$ . We numerically integrate the equations of motion using a fourth-order Runge–Kutta technique with a time step  $dt = T/2,500$ , where  $T = 2\pi/\omega$ .

Each isolated (uncoupled) pendulum operates in a region of parameter space consisting of a region of chaotic orbits separating regions of qualitatively distinct periodic orbits. The default length  $l = 1.0$  results in chaotic motion, characterized by a posi-

<sup>†</sup> Present address: School of Physics, Emory University, Atlanta, Georgia 30322, USA.

tive Lyapunov exponent. Shortening the pendulum results in an overturning motion, a 'rotation', wherein the applied a.c. and d.c. torques combine to rotate the pendulum  $360^\circ$  over the top. Lengthening the pendulum results in a non-overturning motion, a 'libration', wherein the applied torques are insufficient to fully overcome the pendulum's increased rotational inertia.

Figure 1 depicts the continuous spatiotemporal evolution of arrays of  $N=128$  pendula. Time increases from left to right. The colours code the angular velocities of the pendula; red indicates negative velocities and blue positive velocities; the saturation of the colours decreases to zero (white) as the angular speed decreases to zero. Thus narrow bands of red and blue lines represent sudden rapid motion of the arrays. Figure 1a depicts an example of the evolution of a homogeneous array. The predominance of blue over red is a consequence of the d.c. term in the torque. The evolution not only appears disordered but is in fact chaotic: the system is characterized by a positive leading Lyapunov exponent. The subsequent panels depict what happens when the array is disordered by uniformly varying the lengths of the pendula.

Figure 1b depicts one example of introducing a uniform and symmetric disorder of  $\pm 20\%$  so that the pendula lengths are uniformly distributed in the interval  $[0.8, 1.2]$ . After a short transient (not shown), a relatively simple but non-trivial pattern, periodic in space and time, emerges. This pattern repeats every two forcing periods, as indicated by the horizontal black bar. Figure 1c depicts a second example of  $\pm 20\%$  disorder. Operating on a different random realization of lengths, and beginning with a different set of random initial conditions, the array self-organizes a more complex spatiotemporal pattern, one that repeats every 20 forcing periods. Finally, Fig. 1d depicts an example of

introducing a uniform and symmetric disorder of  $\pm 10\%$  so that the pendula lengths are uniformly distributed in the interval  $[0.9, 1.1]$ . This trial resulted in an especially complex pattern. Although the bottom edge of the array repeats every forcing period, other segments repeat every three, four, six and twelve forcing periods, so that the entire array repeats every 24 periods. Through repeated trials we can thus obtain a rich variety of periodic spatiotemporal patterns.

If we disorder the pendula asymmetrically, by either shortening all of them or lengthening all of them, then either rotation or libration dominates the motion. If the lengths are decreased, a fraction of the pendula rotate easily and these quickly entrain the others until the entire array is rotating end-over-end in unison. If the lengths are increased, a fraction of the pendula only librate and these quickly inhibit the others from rotating.

We have obtained similar results in higher dimensions. Figure 2 illustrates a two-dimensional example of a square array of  $128 \times 128$  pendula. Here each pendulum is coupled to its four nearest neighbours. We represent the angular velocities of the square array once for each forcing period. Introducing  $\pm 20\%$  spread in the pendula lengths converts the chaotic motion of the homogeneous array (forcing periods 17–24) into the complex periodic motion of the inhomogeneous array (forcing periods 91–98). Large areas of the array repeat after two forcing periods, whereas some small regions repeat after three or four forcing periods.

We have observed similar phenomena in other extended systems. For example, in numerical experiments (Y.B., J.F.L. and W.L.D., manuscript in preparation), we have used nonuniform realizations of disorder to convert the chaotic motion of a homogeneous array of forced damped Duffing oscillators, described by  $\ddot{x}_n + \gamma \dot{x}_n = kx - k'x^3 + A \sin \omega t + \kappa(x_{n+1} - x_n) + \kappa(x_{n-1} - x_n)$ , into the periodic motion of the corresponding inhomogeneous array. In one example the oscillators of the homogeneous but chaotic array freely explore both wells of the Duffing double-well potential. However, introducing the appropriate inhomogeneity localizes all the oscillators in a single well, thereby regularizing the motion.

We should like to distinguish two elementary mechanisms by which disorder may stabilize a chaotic system. One mechanism requires small but non-zero disorder, whereas the other may involve arbitrarily large disorder. We have observed both these mechanisms at work in the pendulum array and Duffing array in different parameter regimes. (In fact we believe that both mechanisms are involved in the formation of the patterns in Figs 1 and 2). The first mechanism can be best understood by considering the topology of a system's dynamical attractors. An extended system occupies a large many-dimensional parameter space, including all the parameters of the individual components of the system. If the system is dissipative, every neighbourhood in parameter space is typically associated with very many dynamical attractors. Disordering the system by changing its parameters moves it in parameter space and consequently transforms its attractors. Often even a small amount of disorder can shift the dynamics from a chaotic attractor to a nearby periodic attractor. (This topological feature of attractors is often exploited in the control of chaos.) The disorder needed to stabilize a chaotic array depends on both the distance and the direction in parameter space to the nearest periodic attractor. These are controlled by the magnitude and the distribution of the disorder. In our pendulum array, two simple but qualitatively distinct periodic motions, namely synchronized libration and synchronized rotation (both following the external a.c. drive), exist nearby in parameter space and can be reached by lengthening or shortening all the pendula. Many complex nonchaotic motions (attractors), like the ones depicted in Fig. 1 also exist nearby in parameter space and these can be reached by shortening some pendula and lengthening others. For this mechanism to work, it is not essential to remove some or all of the pendula from their chaotic regimes. Indeed, we have succeeded in stabilizing

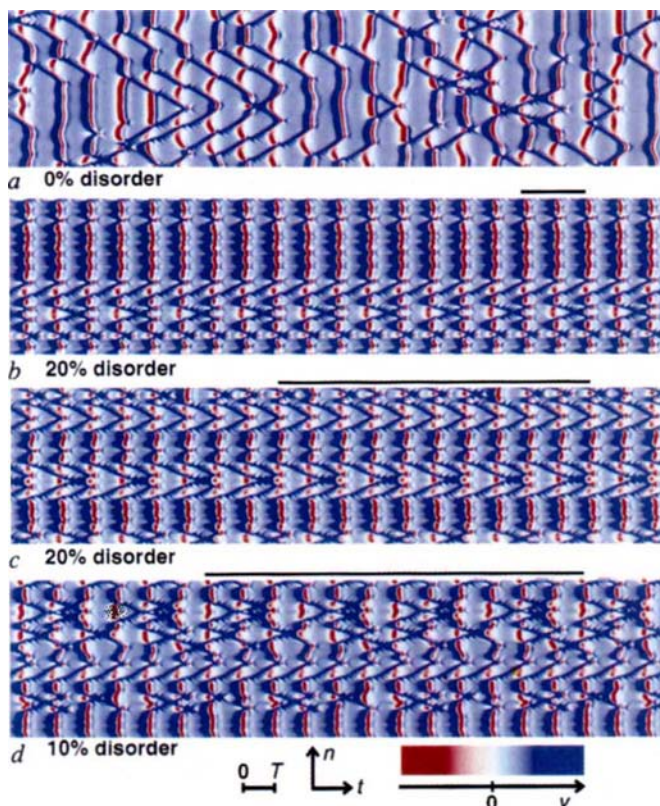
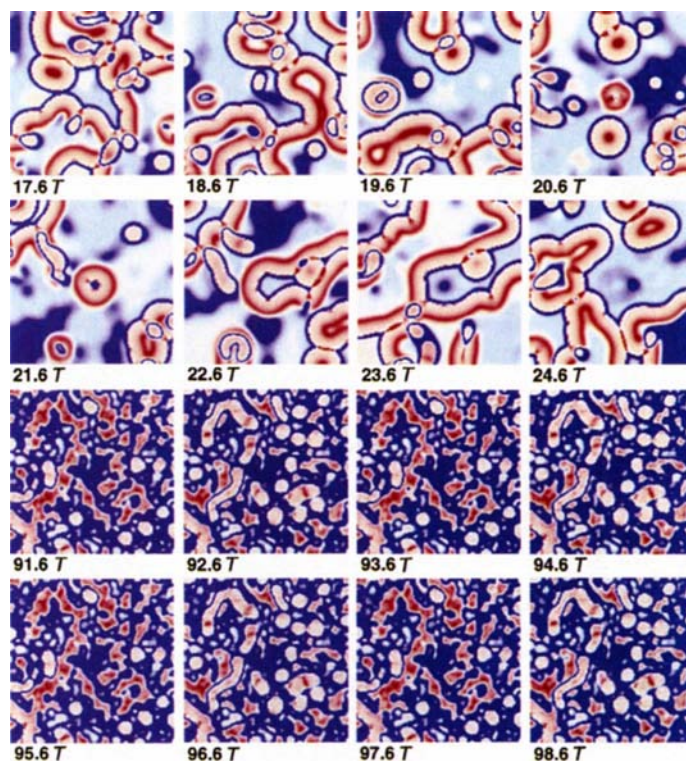


FIG. 1 Spatiotemporal evolution of chains of  $N=128$  coupled pendula. Time increases continuously from left to right. The colours code angular velocities; blue indicates positive velocity and red negative velocities. Horizontal black bars indicate the length (a, b) or half length (c, d) of periodic spatiotemporal patterns. Parameters are:  $\gamma=0.75$ ,  $\tau'=0.7155$ ,  $\tau=0.4$ ,  $\omega=0.25$ ,  $T=25.13$ ,  $\kappa=0.5$ .



FIG. 2 Spatiotemporal evolution of a grid of  $N=128 \times 128$  coupled pendula. The time evolution is 'strobed' once each forcing period  $T$ . Disorder in the grid converts chaotic motion (top) into periodic motion (bottom). The colours and parameters are the same as in Fig. 1.



an array while keeping all the pendula in their chaotic regimes. We obtained this result by creating a simple example of nine different pendula each of which is chaotic when uncoupled but all of which form a periodic array when coupled.

The second simple mechanism by which disorder may stabilize a chaotic array can involve arbitrarily large perturbations. Disorder can be used to remove some of the oscillators of an array from their chaotic band, thereby creating distinctly different populations of oscillators. The periodic population can lock the remaining chaotic population to the external a.c. drive, thereby creating periodic solutions to the equations of motion. (To demonstrate the possibility that coupling stable and unstable maps can yield a stable solution one can analyse the simple linear map  $x_n^{i+1} = \mu_n x_n^i + \kappa(x_{n+1}^i - x_n^i) + \kappa(x_{n-1}^i - x_n^i)$ . The uncoupled maps  $\mu_n = 0$  have fixed points  $x_n = 0$  that are stable if  $\mu_n < 1$  and unstable if  $\mu_n > 1$ . The coupled maps have a stable fixed point  $\mu_n = 0$  even when some  $\mu_n > 1$ .) This second mechanism does not restrict the size of the disorder. Indeed, we found an example, in a different pendulum array, in which 75% uniform disorder optimizes spatiotemporal order by minimizing the leading Lyapunov exponent.

Disorder can tame chaos and foster periodic patterns. We hope that this work will stimulate further study into the regularizing and pattern-forming potential of disorder in extended systems. This phenomenon is possible in situations where arrays of nonlinear elements have a natural or unavoidable spread of features. Our results suggest that the role of disorder in spatially extended systems may be less of a randomizing influence than an intrinsic mechanism of pattern formation, self organization and control.  $\square$

8. Johnson, G. A., Löcher, M. & Hunt, E. R. *Phys. Rev. E* **51**, R1625–R1628 (1995).
9. Garfinkel, A., Spano, M. L., Ditto, W. L. & Weiss, J. N. *Science* **257**, 1230–1235 (1992).
10. Schiff, S. J., Jerger, K., Duong, D. H., Chang, T., Spano, M. L. & Ditto, W. L. *Nature* **370**, 615–620 (1994).
11. Shinbrot, T. A., Grebogi, C., Ott, E. & Yorke, J. A. *Nature* **363**, 411–417 (1993).
12. Ditto, W. L. & Pecora, L. M. *Scient. Am.* **269** (2), 78–84 (1993).

ACKNOWLEDGEMENTS. We thank K. Wiesenfeld for helpful discussions. J.F.L. thanks The College of Wooster for making possible his sabbatical at Georgia Tech.

## Wide-band reflective polarizers from cholesteric polymer networks with a pitch gradient

D. J. Broer, J. Lub & G. N. Mol

Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

CHOLESTERIC liquid crystals, in which the orientation of the molecules varies in a helical fashion, are used for optical filtering of circularly polarized light, for example in liquid-crystal displays. They reflect circularly polarized incident light of the same handedness as the cholesteric helix, and in a wavelength band that depends on the helical pitch (repeat distance). This pitch can be selected by careful design of the liquid-crystal molecules or by mixing cholesteric materials with nematic (linearly oriented) liquid crystals, which tend to increase the pitch. Stable optical filters are produced by crosslinking the cholesteric molecules by photopolymerization<sup>1</sup>; these filters typically have a reflection wavelength bandwidth of  $\sim 50$  nm. Here we show that, by introducing a gradient in the pitch of the cholesteric helix, we can obtain reflection of one of the two circularly polarized components over the entire visible spectrum. Polarizers with such broad-band reflectivity would greatly improve the light yield and energy efficiency of liquid-crystal display devices.

A single-pitch cholesteric liquid crystal reflects light<sup>2</sup> of a wavelength between  $\lambda_1 = pn_o$  and  $\lambda_2 = pn_c$ . Here  $p$  is the pitch of

Received 22 June; accepted 25 October 1995.

1. Wiesenfeld, K. & Moss, F. *Nature* **373**, 33–36 (1995).
2. Lindner, J. F., Meadows, B. K., Ditto, W. L., Inghosa, M. E. & Bulsara, A. R. *Phys. Rev. Lett.* **75**, 3–6 (1995).
3. Rajasekar, S. & Lakshmanan, M. *Physica D* **67**, 282–300 (1993).
4. Sepulchre, J. A. & Babloyantz, A. *Phys. Rev. E* **48**, 945–950 (1993).
5. Gang, H. & Kaifen, H. *Phys. Rev. Lett.* **71**, 3794–3797 (1993).
6. Auerbach, D. *Phys. Rev. Lett.* **72**, 1184–1187 (1994).
7. Aranson, I., Levine, H. & Tsimring, L. *Phys. Rev. Lett.* **72**, 2561–2564 (1994).

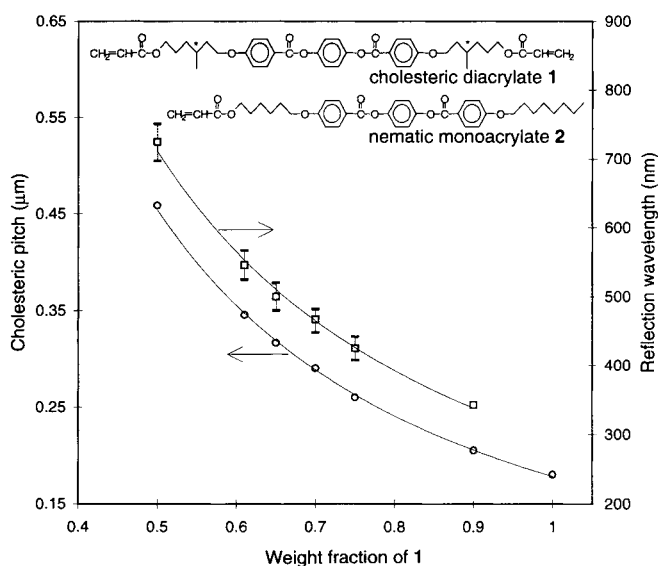


FIG. 1 The pitch of the molecular helix in the cholesteric network and the resulting reflection wavelength as a function of composition. The bars on the reflection data indicate the bandwidth of the reflection. At a wavelength  $<400$  nm, exact detection is hindered by the absorption bands of the materials. The cholesteric pitch at weight fraction 1 of compound **1** is measured by scanning electron microscopy<sup>1</sup>.

the helix corresponding to a  $2\pi$  molecular rotation, and  $n_o$  and  $n_e$  are the ordinary and extraordinary refractive indices of the locally uniaxial structure. Within this reflection band, right-circularly polarized light is reflected by a right-handed helix, whereas left-circularly polarized light is transmitted. Outside the reflection band all polarization states are transmitted. For colourless materials (typically,  $n_o=1.5$  and  $n_e=1.7$ ) the bandwidth in the visible region is  $<75$  nm. By incorporating a gradient in the pitch, we have been able to increase the bandwidth considerably to values larger than 300 nm. The reflection band will then be located between  $\lambda_1=p_1n_o$  and  $\lambda_2=p_2n_e$ , where  $p_1$  and  $p_2$  are the smallest and the largest pitch respectively.

Solid structures are needed to realize a permanent helix gradient. With photo-initiated chain-crosslinking of liquid-crystal monomers it is possible to freeze-in all types of mesophases into stable polymeric structures<sup>3</sup>. Recently, it has been demonstrated that ultraviolet irradiation of the cholesteric diacrylate **1** (Fig. 1) in the presence of a photoinitiator produces a cholesteric polymer network with a pitch of 180 nm. The reflection wavelength of this material is too small (282 nm) but can be shifted to the visible region by blending with a nematic acrylate, for example, liquid-crystal monoacrylate **2**. Figure 1 shows the influence of the composition on the cholesteric pitch and the reflection band. The bandwidth of the samples thus produced is  $\sim 50$  nm.

To initiate the formation of a gradient in the helix, we investigated whether we can set up opposing concentration gradients of the pitch-tightening cholesteric monomer and the pitch-widening nematic monomer during polymerization, by diffusion. The difference in reactivity between diacrylate **1**, having two reactive sites per molecule, and monoacrylate **2**, having only one reactive site, provides a means to achieve this. By adding a small amount of dye **3** (Fig. 2), with an absorption maximum at 334 nm, close to that of the photoinitiator (Irgacure 651, Ciba-Geigy; 345 nm) and an extinction coefficient two orders larger than that of the photoinitiator, a gradient in ultraviolet intensity can be achieved over the thickness of the sample. The polymerization rate will then be fastest at the top of the film (lamp-side) resulting in a faster consumption of the most reactive monomer at the same

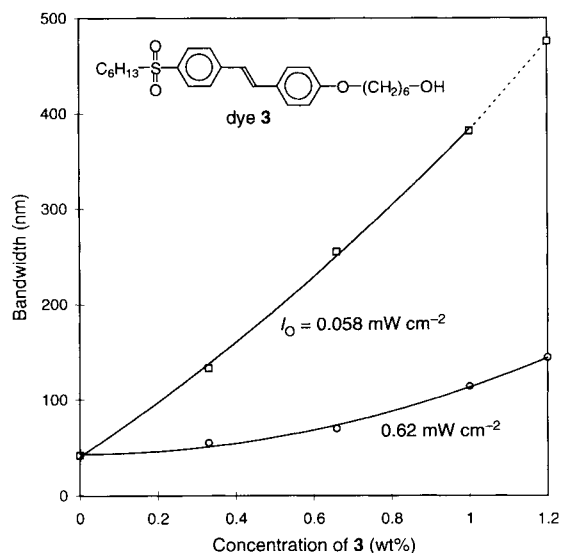


FIG. 2 Bandwidth of the cholesteric network as a function of the concentration of added dye **3**. Monomer composition is **1/2**=0.6/0.4, corresponding to a centre position of the reflection band at around 546 nm. The source of ultraviolet light for polymerization was a Philips PL10 fluorescent lamp emitting at 365 nm. The intensity  $I_0$  at the location of the sample was  $0.62 \text{ mW cm}^{-2}$  ( $\circ$ ) or, by using a 10% grey filter,  $0.058 \text{ mW cm}^{-2}$  ( $\square$ ). (Before polymerization the bandwidth is independent of the concentration of **3**, and is between 40 and 50 nm. The inaccuracy for bandwidths  $>400$  nm is due to absorption of the samples in the ultraviolet region.)

location. The depletion of this monomer starts a diffusion process in which the upper part of the layer is enriched with cholesteric material, and the lower part with nematic material.

Figure 2 shows the influence of the dye concentration on the bandwidth. Already with 1 wt% dye the bandwidth can be increased from 42 to 381 nm. This implies that for this specific sample the cholesteric pitch increases from 0.23  $\mu\text{m}$  at the top to 0.44  $\mu\text{m}$  at the bottom of the film. Scanning electron microscopy (Fig. 3) showed that the pitch increases linearly over the thickness. As diffusion during polymerization is the key mecha-

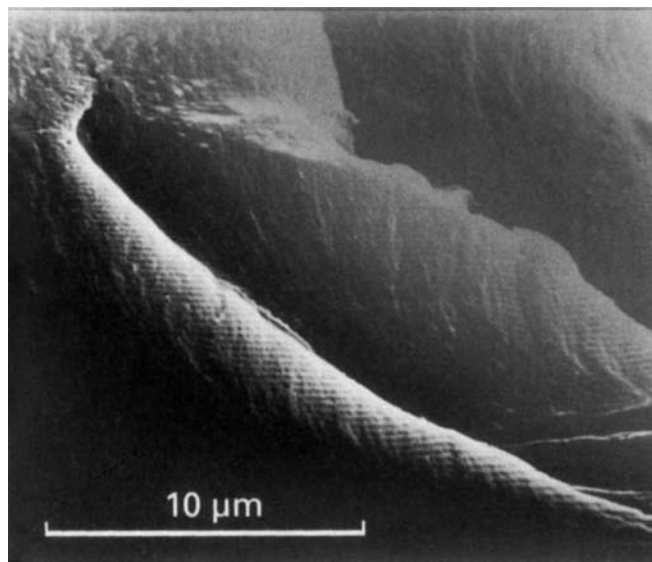


FIG. 3 Scanning electron microscopy image of the fracture surface of a pitch-gradient cholesteric network. The dimension of the bands in the fine structure, displayed on the fracture planes, corresponds to a  $\pi$  molecular rotation, that is,  $p/2$ . The bands increase linearly in thickness from the bottom to the top of this cross-section of the film.

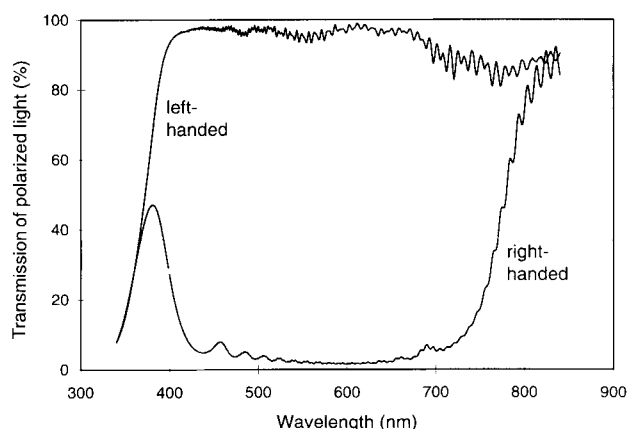


FIG. 4 Transmission of circularly polarized light of a 15- $\mu\text{m}$ -thick film of a cholesteric network with pitch gradient, as a function of wavelength. The results for left- and right-handed circularly polarized light are shown. The calculated extremes of the pitch are 0.27  $\mu\text{m}$  at the top, and 0.45  $\mu\text{m}$  at the bottom, of the film.

nism for the formation of the pitch gradient, the polymerization rate should be balanced to the monomer diffusion rate. Decreasing the rate of polymerization, for instance by decreasing the ultraviolet light intensity, increases the bandwidth because it permits a more extensive monomer diffusion before immobilization by the network occurs. Increasing the polymerization temperature affects predominantly the diffusion rate and therefore leads to broader reflection bands. The dye concentration increases the intensity gradient (and thus the driving force for diffusion) and decreases the integral polymerization, yielding the progressive increase of the bandwidth with dye concentration.

The pitch-gradient cholesteric networks are clear, colourless and have a metallic appearance as they reflect 50% of the light evenly over the visible spectrum. Figure 4 shows the polarization selectivity of the reflection band measured by polarized ultraviolet-visible spectroscopy. The reflection of right-handed circularly polarized light of this film extends over a bandwidth of 350 nm, allowing only 2% transmission in the visible region, whereas left-handed circularly polarized light at wavelengths above 400 nm is completely transmitted. The films are well suited as reflective polarizers. The bandwidth is variable, and the band position can be located in the range from ultraviolet to the infrared by changing composition and polymerization conditions. The reflective polarizer is selective for circularly polarized light which, if desired, can be easily converted into linearly polarized light by adhering to the cholesteric layer a quarter-wave retardation foil.

We have used these materials in prototype flat-panel LCDs. Inserting a stack of the pitch-gradient cholesteric layer and the quarter-wave foil between the back-light system and the dichroic polarizer improves the light yield by 40% without any optimization of the back-light system. Wrongly-polarized light which is reflected from the cholesteric layer can be recycled in the back-light system, as has been shown for narrow-band cholesterics in a projection LCD set-up<sup>4</sup>. For use in a direct-view flat-panel LCD, a diffusor foil can be placed between the polarizer and the back-light. The light reflected from the cholesteric polarizer then becomes depolarized by multiple scattering in the diffusor film, and is reflected into the viewer's direction by the diffusor foil and by the reflector in the back-light system. As this process of polarization by reflection, depolarization and back-reflection may repeat itself, high yields of polarized light are possible as long as absorption is suppressed.  $\square$

- Broer, D. J. *Radiation Curing in Polymer Science and Technology* Vol. 3, Ch. 12, 383–443 (Elsevier Science, London, 1993).
- Belayev, S. V. et al. *Jap. J. appl. Phys.* **29**, L634–L637 (1990).

ACKNOWLEDGEMENTS. We thank J. H. van der Veen and C. M. R. de Witz for their contribution to the synthesis of monomers and realizing prototypes of the reflective polarizers respectively.

## Self-assembly of ten molecules into nanometre-sized organic host frameworks

Makoto Fujita\*, Daichi Oguro\*,  
Mayumi Miyazawa\*, Hiroko Oka\*,  
Kentaro Yamaguchi† & Katsuyuki Ogura\*

\* Department of Applied Chemistry, Faculty of Engineering and  
† Chemical Analysis Center, Chiba University, Yayoicho, Inageku,  
Chiba 263, Japan

THE synthesis of hollow, nanometre-scale molecular 'container compounds'<sup>1,2</sup> makes possible the creation of localized chemical micro-environments with properties different from those of the bulk phases; such compounds can be used, for example, to encapsulate otherwise unstable molecular species<sup>3</sup>. Container compounds have previously been prepared by conventional chemical synthesis<sup>1,2</sup>. Here we report the construction of a hollow, roughly spherical supramolecular framework by self-assembly<sup>4–8</sup>. The framework, which is ~2–5 nm in diameter, is constructed from ten species: four organic ligands held together by six metal ions. It has tetrahedral symmetry, and has a large central void, in which guest molecules can be accommodated. We show that four adamantyl carboxylate ions can be encapsulated within this self-assembling cage.

Coordination of ligands containing diverging 4-pyridyl groups to Pd(II) complexes like **1** (Fig. 1), which possess two vacant *cis* coordination sites, gives rise to the assembly of well defined structures such as macrocycles<sup>9</sup>, cages<sup>12</sup> and catenanes<sup>13,14</sup>. When **1** was combined with the rigid tridentate ligand **2a** (Fig. 1) in a 3:2 ratio in water, we observed (by <sup>1</sup>H NMR) the quantitative self-assembly of a single component (Fig. 2a). The 4-pyridyl groups or ethylenediamine moieties appear identical by spectroscopy, and give the component its symmetry. The thermodynamic stability of this product is remarkable: when the components **1** and **2a** were mixed in a 4:2 ratio, only the assembly of the 3:2 product was again observed, and excess free **1** remained intact (Fig. 2b). This stable 3:2 product must be an adamantane-like three-dimensional cage<sup>15,16</sup> compound, shown as **3a** in Fig. 1a, having a (1)<sub>6</sub>(2)<sub>4</sub> composition because only this structure is consistent with the observed high symmetry and thermodynamic stability. By simply adding ethanol to the aqueous solution, **3a** was precipitated as a pure form and isolated in 90% yield; the elemental analysis was consistent with a formula **3a** · 17H<sub>2</sub>O · 2C<sub>2</sub>H<sub>5</sub>OH.

The structure of **3a** was confirmed by X-ray crystallographic analysis of its clathrate complex with four adamantyl carboxylate ions (**4**) (Fig. 1b). Crystals were grown by allowing an aqueous solution of **3a** in the presence of **4** · Na<sup>+</sup> to stand at 20 °C for three weeks. Elemental analysis of the caged guest shows that it has eight NO<sub>3</sub><sup>-</sup> ions, indicating that it exists in the free anion form, and not as an acid (**4** · H<sup>+</sup>) or sodium salt (**4** · Na<sup>+</sup>). Thus, the formula of the clathrate complex is represented as **3a** · (**4**)<sub>4</sub> (**3a**' = **3a** – NO<sub>3</sub><sup>-</sup>)<sub>4</sub>. The inclusion geometry of guest **4** is very interesting: hydrophobic adamantyl groups are located inside while hydrophilic carboxylate groups are outside, and the main axes of the four guest molecules point to the corners of a tetrahedron. The bulky guest molecule seems to stabilize the crystal by filling the large void of **3a**' which otherwise would contain large amounts of disordered solvent molecules. In fact, although crystals were also obtained in the absence of **4**, they

Received 14 August; accepted 10 October 1995.

- Lub, J., Broer, D. J., Hikmet, R. A. M. & Nierop, K. G. J. *Liq. Cryst.* **18**, 319–326 (1995).
- Oseen, C. W. *Trans. Faraday Soc.* **29**, 883–899 (1993).

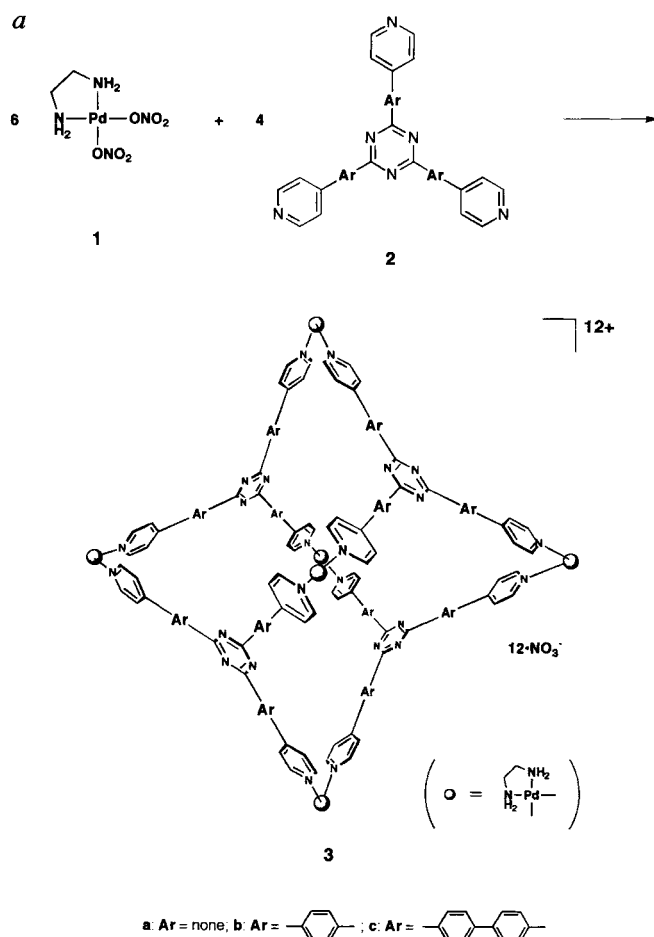


FIG. 1 **a**, The reaction scheme for the self-assembly of compound **3**. **b**, The crystal structure of the clathrate complex **3a'·(4)<sub>4</sub>** (a space-filling model presentation). A crystal with dimensions  $0.42 \times 0.40 \times 0.34$  mm was sealed in a glass capillary. A laser-stimulated fluorescence image plate was used as a two-dimensional area detector. Because of the instability of the crystal, rapid analysis was required. Thus, 15 frames were recorded at intervals of  $6^\circ$ , and each exposure lasted for 5 min;  $\sim 105$  min was required for the total data collection. The distance between the crystal and the detector was 100 mm. Accurate cell parameters were independently measured on a four-circle diffractometer. Crystal data: tetragonal,  $I4_1/amd$  (no. 141);  $a = 26.913(3)$  Å,  $c = 27.383(6)$  Å;  $V = 19,842(2)$  Å<sup>3</sup>;  $z = 4$ ;  $F(000) = 7,568$ ;  $\mu(\text{Mo K}\alpha) = 48.11 \text{ cm}^{-1}$ ;  $\lambda(\text{Mo K}\alpha) = 0.71070$  Å; temperature  $-100^\circ\text{C}$ ; 3,413 reflections measured, 1,180 observed ( $I > 4.50\sigma(I)$ ); number of parameters 251;  $R = 0.125$ ;  $R_w = 0.164$ . (Crystallographic data are available: see Supplementary Information.)

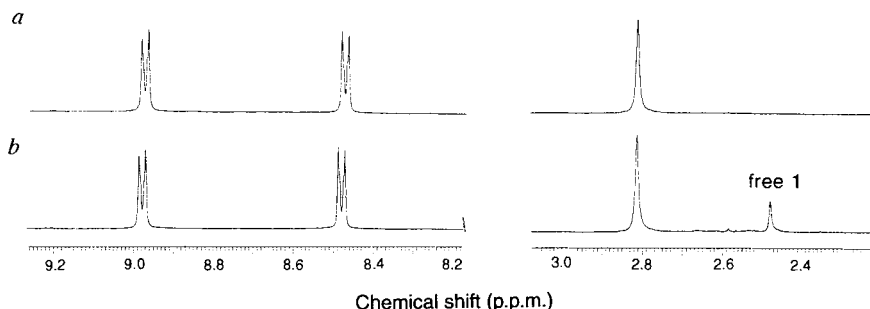
were too unstable for diffraction studies. Though 210 heavy atoms are involved in **3a'·(4)<sub>4</sub>**, the crystallographic molecule contains only 29 independent atoms, which give a highly symmetric structure after symmetry operations (Fig. 1b). Neither distorted bond lengths nor angles were found in the framework of **3a'**, which is consistent with the supermolecule's stability. In particular, N(pyridine)–Pd–N(pyridine) bond angles are almost ideal right-angles ( $88.5\text{--}90.4^\circ$ ). In spite of the data collection at  $-100^\circ\text{C}$ , nitrate ions are highly disordered and missing in the crystallography.

The encapsulation of four molecules of **4** by supramolecule **3a'** was also observed in an aqueous solution. The titration of **3a** with guest **4** was monitored by  $^1\text{H}$  NMR spectroscopy (Fig. 3). In the spectrum of **3a** and **4** at the 1:1 stoichiometry, 75% of **3a** remained intact while 25% appeared as an independent set of signals; a very-high-upfield shift of **4** was observed

concurrently (Fig. 3a). This observation indicates the formation of the **(3a')·(4)<sub>4</sub>** complex, as well as the slow exchange between free and complexed **3a** on the NMR timescale. At the 1:2 stoichiometry, 50% of **3a** was complexed (Fig. 3b); finally, only **(3a')·(4)<sub>4</sub>** complex was observed at the 1:4 stoichiometry (Fig. 3c). Furthermore, at the 1:8 stoichiometry, excess **4** appeared as a free form (Fig. 3d). The same inclusion geometry of the guest observed in the crystal structure was suggested in the aqueous case by the more marked upfield shift of guest protons at the hydrophobic site. The absence of the intermediary **(3a')·(4)<sub>n</sub>** ( $n = 1\text{--}3$ ) is explained in terms of allosteric effects; that is, the complexation between **3a** and **4** becomes more effective with increasing number of the encapsulated molecules because of enhanced hydrophobicity in the cavity.

The framework of **3a** can be further expanded by employing phenylene homologues of **2a**. From **1** and ligand **2b** in

FIG. 2  $^1\text{H}$  NMR spectra (400 MHz,  $\text{D}_2\text{O}$ ) of the self-assembled product from the components **1** and **2** which were combined at a 3:2 or 4:2 stoichiometry. **a**, At the 3:2 stoichiometry. Two doublet-like signals appearing at chemical shift ( $\delta$ ) 8.98 and 8.48 p.p.m. are assigned to the PyHa and PyH $\beta$ , respectively, of **3a**. A singlet signal at  $\delta$  2.81 p.p.m. is assigned to  $\text{CH}_2\text{CH}_2$  protons of **3a**. **b**, At the 4:2 stoichiometry. A singlet signal newly appearing at  $\delta$  2.48 p.p.m. is assigned to  $\text{CH}_2\text{CH}_2$  protons of unreacted **1**.





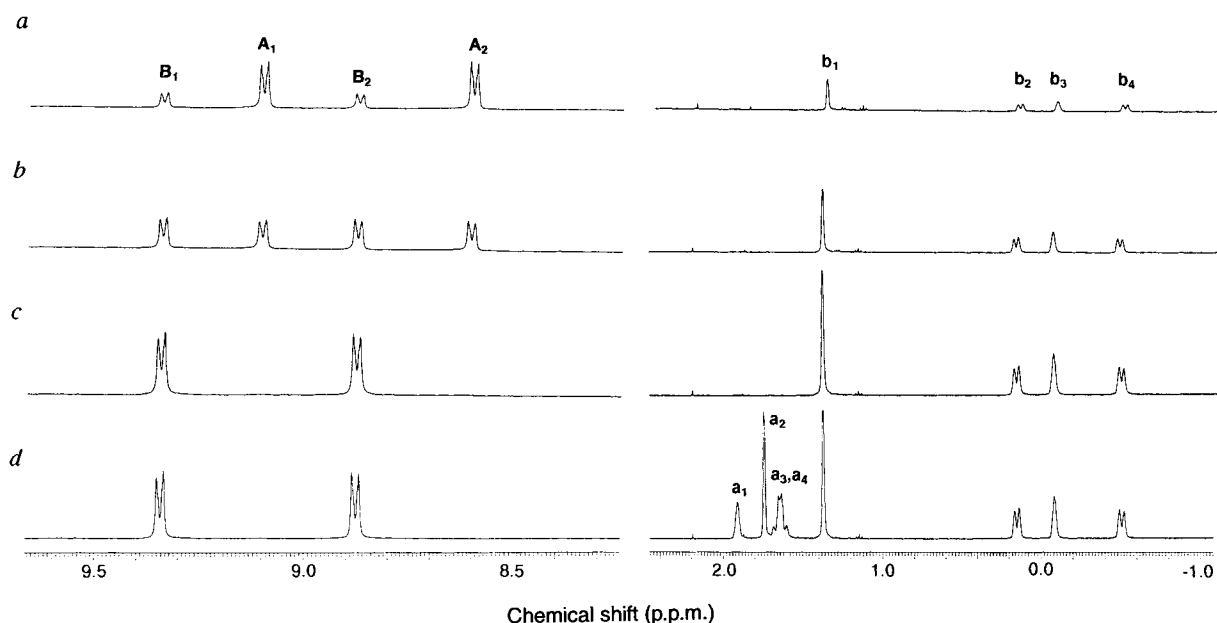


FIG. 3 Monitoring of the titration of **3a** with **4** by  $^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ , external TMS). The ratios **3a**:**4** are: a, 1:1; b, 1:2; c, 1:4; d, 1:8. Components (signals) are as follows: uncomplexed **3a** (peaks  $A_1$  and

$A_2$ ); complexed **3a** (peaks  $B_1$  and  $B_2$ ); uncomplexed **4** (peaks  $a_1$ ,  $a_2$ ,  $a_3$  and  $a_4$ ); complexed **4** (peaks  $b_1$ ,  $b_2$ ,  $b_3$  and  $b_4$ ).

3:2 stoichiometry, we again observed the self-assembly of a single and highly symmetric product assigned as phenylene homologue **3b** (Fig. 4a), which was isolated with 73% yield. Similarly, the bisphenylene homologue **3c** was also assembled in a high yield from **1** and the corresponding tridentate ligand **2c** (Fig. 4b).

Supramolecules **3a–c** belong to the category of ultrafine particles because their sizes reach to the nanometre scale (2–5 nm). The crystallography, as well as molecular modelling, showed that the largest intramolecular Pd–Pd distances are 1.9, 3.3 and 4.6 nm in **3a**, **3b** and **3c**, respectively. The particulate characteristic of these supramolecules was demonstrated by volume-weighted distribution analysis of the particle sizes by a laser light

scattering method. (Original data available, see Supplementary Information.) The supramolecules **3a** and **3b** were shown to have mean diameters of 1.5 and 4.6 nm, respectively.

Recently, Robson and co-workers reported<sup>17,18</sup> novel assemblies of transition metals and ligand **2a** into three-dimensional infinite frameworks that possess large voids. We emphasize that, in our work, the employment of ethylenediamine as a terminating group on Pd atoms makes it possible to limit their assembly to the nanometre-scale to the molecular level, giving molecular-based 'ultrafine particles'.<sup>19</sup>

Received 6 June; accepted 23 October 1995.

1. Cram, D. J. *Nature* **356**, 29–36 (1992).
2. Cram, D. J. & Cram, J. H. *Container Molecules and Their Guests* (Royal Society of Chemistry, Cambridge, 1994).
3. Cram, D. J., Tanner, M. E. & Thomas, R. *Angew. Chem., int. Edn engl.* **30**, 1024–1027 (1991).
4. Lehn, J.-M., Rigault, A., Siegel, J., Harrowfield, J. & Chevrier, B. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2565–2569 (1987).
5. Lehn, J.-M. *Angew. Chem., int. Edn engl.* **29**, 1304–1319 (1990).
6. Fabbrizzi, L. & Poggi, A. (eds) *Transition Metals in Supramolecular Chemistry* (Kluwer, Dordrecht, 1994).
7. Baxter, P., Lehn, J.-M. & DeCian, A. *Angew. Chem., int. Edn engl.* **32**, 69–72 (1993).
8. Baxter, P. N. W., Lehn, J.-M., Fischer, J. & Youinou, M.-T. *Angew. Chem., int. Edn engl.* **33**, 2284–2287 (1994).
9. Fujita, M., Yazaki, J. & Ogura, K. *J. Am. chem. Soc.* **112**, 5645–5647 (1990).
10. Fujita, M., Nagao, S., Iida, M., Ogata, K. & Ogura, K. *J. Am. chem. Soc.* **115**, 1574–1576 (1993).
11. Rauter, H., Hillger, E. C., Exleben, A. & Lippert, B. *J. Am. chem. Soc.* **116**, 616–624 (1994).
12. Fujita, M., Nagao, S. & Ogura, K. *J. Am. chem. Soc.* **117**, 1649–1650 (1995).
13. Fujita, M., Ibukuro, F., Hagihara, H. & Ogura, K. *Nature* **367**, 720–724 (1994).
14. Fujita, M., Ibukuro, F., Yamaguchi, K. & Ogura, K. *J. Am. chem. Soc.* **117**, 4175–4176 (1995).
15. Saalfrank, R. W., Stark, A., Peters, K. & Schnering, H. G. v. *Angew. Chem., int. Edn engl.* **27**, 851–853 (1988).
16. Saalfrank, R. W. et al. *Angew. Chem., int. Edn engl.* **33**, 1621–1624 (1994).
17. Batten, S. R., Hoskins, B. F. & Robson, R. *Angew. Chem., int. Edn engl.* **34**, 820–822 (1995).
18. Batten, S. R., Hoskins, B. F. & Robson, R. *J. Am. chem. Soc.* **117**, 5385–5386 (1995).
19. Knapen, J. W. et al. *Nature* **372**, 659–663 (1994).

SUPPLEMENTARY INFORMATION. Requests should be addressed to Mary Sheehan at the London editorial office of *Nature*.

ACKNOWLEDGEMENTS. We thank Nippon Kayaku Co. Ltd and Nozaki Sangyo Co. Ltd. for the distribution analysis of particle sizes. This work was supported by the Ministry of Education, Science and Culture, Japan.

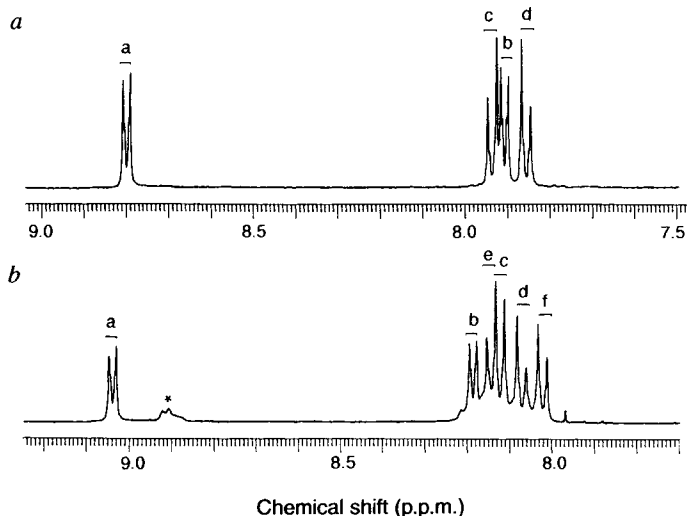


FIG. 4 a,  $^1\text{H}$  NMR spectra (400 MHz,  $\text{D}_2\text{O}$ ) of **3b**. Assignments: a,  $\text{PyH}_4$ ; b,  $\text{PyH}\beta$ ; c and d,  $\text{C}_6\text{H}_4$ . b,  $^1\text{H}$  NMR spectra (400 MHz,  $\text{D}_2\text{O}$ ) of **3c**. Assignments: a,  $\text{PyH}_4$ ; b,  $\text{PyH}\beta$ ; c and d,  $\text{C}_6\text{H}_4$ ; e and f,  $\text{C}_6\text{H}_4$ ; an asterisk, impurities.

# Control of protein–ligand recognition using a stimuli-responsive polymer

Patrick S. Stayton\*, Tsuyoshi Shimoboji†‡, Cynthia Long\*, Ashutosh Chilkoti\*, Guohua Chen†, J. Milton Harris§ & Allan S. Hoffman†

\*† Center for Bioengineering, \* Box 35-7962, † Box 35-2255, University of Washington, Seattle, Washington 98195, USA

§ Department of Chemistry, University of Alabama at Huntsville, Huntsville, Alabama 35899 USA

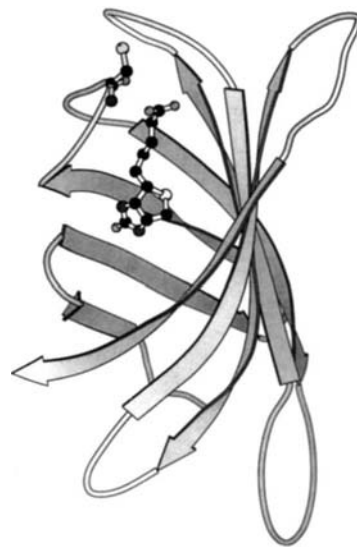
**STIMULI-responsive polymers exhibit reversible phase changes in response to changes in environmental factors such as pH or temperature<sup>1–14</sup>. Conjugating such polymers to antibodies and proteins provides molecular systems for applications such as affinity separations, immunoassays and enzyme recovery and recycling<sup>15–25</sup>. Here we show that conjugating a temperature-sensitive polymer to a genetically engineered site on a protein allows the protein's ligand binding affinity to be controlled. We synthesized a mutant of the protein streptavidin to enable site-specific conjugation of the responsive polymer near the protein's binding site. Normal binding of biotin to the modified protein occurs below 32 °C, whereas above this temperature the polymer collapses and blocks binding. The collapse of the polymer and thus the enabling and disabling of binding, is reversible. Such environmentally triggered control of binding may find many applications in biotechnology and biomedicine, such as the control of enzyme reaction rates and of biosensor activity, and the controlled release of drugs.**

We have described<sup>25</sup> the site-specific conjugation of a temperature-sensitive polymer to a genetically engineered protein. A single thiol group was introduced into cytochrome *b*<sub>5</sub>, which lacks native cysteine thiol groups, in order to precisely position the conjugation site<sup>26</sup>, and a poly(*N*-isopropylacrylamide) chain with one thiol-reactive maleimide end group was conjugated to that specific site at a strict ratio of one polymer chain per protein molecule. That reaction site, however, was intentionally engineered away from the electron-transfer site of the cytochrome to preserve electron-transfer activity even when the conjugate was thermally phase-separated from solution by raising the temperature above 32 °C, the collapse temperature of poly(*N*-isopropylacrylamide)<sup>2</sup>.

In the present study we have used cassette site-directed mutagenesis to create the N49C mutant streptavidin, in which a polymer conjugation site is near the outer edge of the biotin-binding pocket (Fig. 1). Cys was substituted for Asn at position 49, which is hydrogen-bonded to the terminal carboxylate of the valeric acid hydrocarbon side-chain of biotin<sup>27</sup>. To enhance the thiol-reactivity of the poly(*N*-isopropylacrylamide) (hereafter referred to as poly(NIPAAm)), we synthesized a vinyl sulphone terminal derivative (Fig. 2). The conjugation of the vinyl sulphone-poly(NIPAAm) to the N49C streptavidin was performed in solution at pH 7.0 in phosphate-buffered saline, with Tris(2-carboxyethyl)phosphine as a disulphide reducing agent<sup>28</sup>, using a 50 : 1 molar excess of polymer to protein. The polymer–protein conjugate was separated from unreacted protein by thermal precipitation of the poly(NIPAAm) at 37 °C, which phase-separates the streptavidin conjugate. Characterization by high-performance liquid chromatography using a gel permeation chromatography sizing column and polyacrylamide gel electrophoresis (SDS-PAGE) indicated that all streptavidin purified in this manner was conjugated to poly(NIPAAm).

The poly(NIPAAm)–streptavidin conjugate was then immobilized on a biotinylated microporous membrane. Hydrophilic poly(vinylidene difluoride) membranes (Millipore, Bedford, MA) was first coated with polyethyleneimine (*M*<sub>n</sub> 50,000), an amine-containing polymer, by physical adsorption. A biotinylated surface was then prepared by conjugating the *N*-hydroxy succinimidyl ester of biotin to the amine groups of the polyethyleneimine. The N49C streptavidin–poly(NIPAAm) conjugate was adsorbed by affinity binding of the streptavidin to the biotinylated membrane. For control experiments, wild-type streptavidin and the N49C streptavidin mutant were also adsorbed onto the biotinylated membranes, at comparable loading densities to the conjugate. At 4 °C the poly(NIPAAm) is below its phase transition temperature (32 °C), and the polymer adopts a hydrated coil conformation. At this temperature, the biotin-binding capacity of the N49C streptavidin conjugate on the membrane was essentially identical to the wild-type or N49C mutant streptavidin controls. However, a sharp decrease in the biotin-binding capacity of the conjugate on the membrane was observed when the temperature was raised to 37 °C, where the polymer is a collapsed globule (Fig. 3). The wild-type and N49C streptavidin controls did not show temperature-dependent biotin binding capacity. Figure 4 summarizes these results, showing an 84% drop in biotin association for the conjugate at 37 °C relative to 4 °C, whereas the wild-type and N49C streptavidin controls each exhibit only ~2% drop.

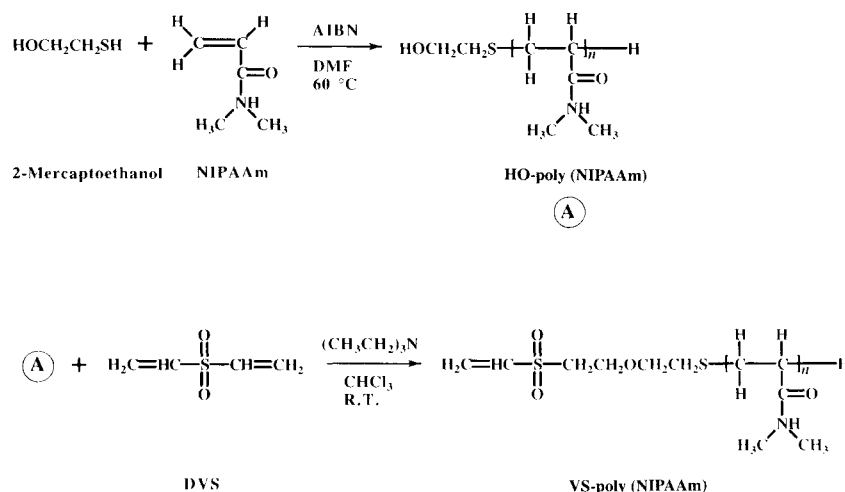
Further evidence supporting a direct link between the control of biotin binding and the temperature-responsive phase-change behaviour of the conjugated poly(NIPAAm) was obtained by measuring the relative percentage blockage of biotin association as a function of temperature between 4 °C and 37 °C. Figure 5 shows the effect of temperature on the ability of the poly



**FIG. 1** Three-dimensional model of the streptavidin subunit with a modelled cysteine at amino acid position 49, showing the spatial relationship between the biotin-binding site and the polymer conjugation site (MOLSCRIPT; ref. 30). The N49C site-directed streptavidin mutant was constructed by cassette mutagenesis techniques, using a synthetic 'core' streptavidin gene designed for bacterial expression<sup>29</sup>. The streptavidin gene was cut between the *Kpn*I and *Xba*I sites, and a synthetic cassette introducing the desired Asn to Cys codon change was subsequently ligated between these sites. The N49C streptavidin mutant sequence was confirmed by automated dideoxy sequencing and subcloned into pET 11 (Novagen, Madison, WI) for expression in *E. coli* strain BL21(DE3). The overexpressed protein was refolded using established protocols<sup>29</sup>, with the addition of dithiothreitol (DTT) (0.5 mM) to maintain the thiol oxidation site.

‡ Mitsubishi Chemical Corporation, Research and Development Division, Yokohama Research Center, 1000, Kamoshida-cho, Aoba-ku, Yokohama 227, Japan.

FIG. 2 Synthesis of the vinyl sulphone terminal derivative of poly(NIPAAm) (VS-poly(NIPAAm)). The hydroxyl terminated poly(NIPAAm) was synthesized by a chain transfer free radical polymerization using a free radical initiator, azobisisobutyronitrile (AIBN), and a chain transfer agent, 2-mercaptoethanol (2ME), in the molar ratio of (NIPAAm)/2ME/AIBN=100/6/1. A 5 mol% excess of divinylsulphone (DVS) and a 5 mol% excess of triethylamine were reacted with the hydroxyl-terminated poly(NIPAAm) (number average weight,  $M_n$ =3,800) in chloroform for 24 h at room temperature. The product was recovered by precipitation with ethyl ether. The product showed characteristic vinyl sulphone peaks in the proton NMR spectrum (in  $d_6$ -DMSO) at 6.21 p.p.m. (two hydrogens) and 6.97 p.p.m. (one hydrogen).



(NIPAAm) chain conjugated to streptavidin to block the biotin binding site as it collapses (open circles), and the phase separation of free poly(NIPAAm) in identical buffer conditions as measured by decreasing light transmission (filled diamonds). Both curves decrease across the same temperature range, with nearly identical midpoints. The initiation of blocking activity of the conjugate precedes the sharp cloud-point transition of the free polymer, which probably reflects the initiation of the gradual collapse of individual poly(NIPAAm) chains at the biotin binding pockets. The free poly(NIPAAm) chains in solution, on the other hand, must reach a critical level of hydrophobicity as they release bound water, and then they must collide, in order to aggregate together and precipitate. This leads to a cooperative phase separation and the observed sharp reduction in light transmission. Thus the gradual collapse of the poly(NIPAAm) chains precedes the critical aggregation point and can be 'visualized' in the activity assay for the conjugate. This correlation of biotin binding to streptavidin with the expansion or collapse of the poly(NIPAAm) random coil indicates that the affinity recognition process is indeed being controlled by precise temperature control of the size of the polymer coil.

The site-specific conjugation of such phase reversible polymers to genetically engineered proteins should provide very sensitive environmental control of biomolecular recognition processes, allowing one to 'gate' on-off rates as well as to 'switch' ligands on or off the receptor protein binding pocket. By varying the relative molecular mass (and thus the polymer coil size) and the composition of the stimuli-responsive polymer, one can vary the type and magnitude of the stimulus needed to trigger a response. Therefore, polymer-protein conjugates can be engineered for particular applications. Furthermore, the conjugate may be used in solution, or immobilized at an interface, on a surface or within a hydrogel.

We envisage many different applications of such molecular gates and switches which could have widespread utility in medicine, biotechnology and information technologies. For example, control of the polymer coil size within or near the active site of a polymer-enzyme conjugate would permit control of enzyme-substrate reaction rates, and could also be selective for reaction with smaller substrates, owing to 'gating' (blocking) of large substrates. Similarly, mixtures of peptide and protein ligands of different sizes could be simultaneously separated by both affinity and size. These actions could be useful in fermentation and bioprocessing technologies in the food and pharmaceutical industries, and in affinity separations of all types. Depending on where the polymer is conjugated to the enzyme or affinity protein, its reversible phase change might also stimulate allosteric modification of the active site, and result in allosteric regulation of reaction or recognition processes.

Triggered release ('switching') of bound ligands could be used to release therapeutics for topical drug delivery to the skin or mucosal surfaces of the body, or for localized delivery of drugs within the body by stimulated release at pre-targeted sites using optical-fibre catheters. Triggered release could also be used to release and recover affinity-bound ligands from chromatographic and other supports in eluate-free conditions, including capture and release of specific cell populations to be used in stem-cell and bone-marrow transplantation. For delicate target ligands such as peptides and proteins, recovery could be affected without the need for time-consuming and harsh elution conditions. Triggered release could also be used to remove inhibitors, toxins or fouling agents from the recognition sites of immobilized or free enzymes and affinity molecules, such as those used in biosensors, diagnostic assays or affinity separations. This could be used to 'regenerate' such recognition proteins for extended process use.

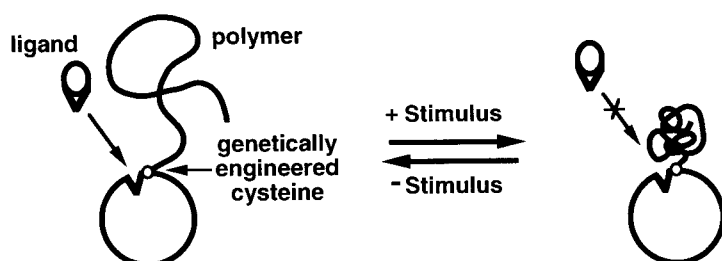


FIG. 3 Schematic illustration of the molecular 'gate' created by conjugating a stimuli-responsive polymer to a genetically engineered site near the receptor binding pocket (or active site) of a protein, which is then immobilized on a solid support. In the hydrated, random coil state, the polymer interferes minimally with ligand binding to the receptor binding pocket. When the phase transition is triggered by an environmental stimulus, such as a small change in temperature, the collapsed polymer can block access and reversibly 'gate' the ligand association.

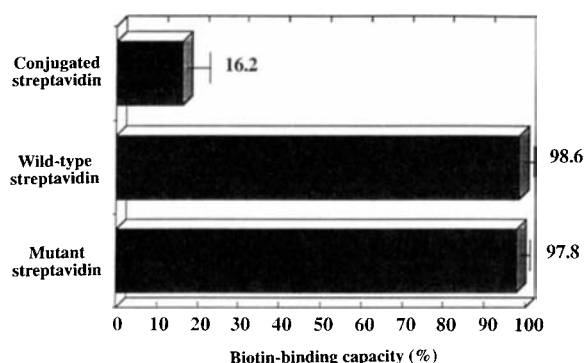


FIG. 4 Blocking of biotin binding to the N49C streptavidin-polymer conjugate by poly(NIPAAm), on a microporous membrane. The histograms represent the percentage biotin-binding capacity of  $^3\text{H}$  biotin (8,9- $^3\text{H}$  biotin, NEN/Dupont) at 37 °C relative to 4 °C for the N49C-polymer conjugate, compared to the wild-type and N49C streptavidin controls. The respective streptavidin proteins were adsorbed onto the biotinylated membranes in 100 mM phosphate buffer (pH 7.0) at similar loading densities, as quantified by  $^3\text{H}$  biotin binding to the adsorbed streptavidin surfaces. The  $^3\text{H}$  biotin-binding capacity of the membranes were then measured in triplicates at 4 °C and 37 °C.

Synthesis of micro- and nano-patterned surfaces of immobilized stimuli-responsive-engineered protein conjugates, and controlled assembly of subsequent layers should be possible with the use of photo-microlithographic screens and other lithographic techniques. The immobilized conjugates could be used to immobilize other biomolecular targets, such as peptide and oligonucleotide libraries, in specific patterns depending on the lithographic generation of free binding sites and subsequent immobilization of the target molecules. Similarly, spatially addressable polymer-protein assemblies could find applications in electro-optical information storage and retrieval by stimulat-

ing the reversible alteration of chromophore or prosthetic group properties near the site of polymer collapse. □

Received 21 June; accepted 20 October 1995.

- Hoffman, A. S. *Artif. Organs* **19**, 458-467 (1995).
- Heskins, M. & Guillet, J. E. *J. Macromol. Sci. Chem.* **A2(8)**, 1441-1455 (1968).
- Tanaka, T. *Scient. Am.* **244**, 124-138 (1981).
- Cussler, E. L., Stokar, M. R. & Varberg, J. E. *Am. Inst. Chem. Eng. J.* **30**, 578-582 (1984).
- Ishihara, K., Hanada, N., Kato, S. & Shinohara, I. *Polymer J.* **16**, 625-631 (1984).
- Hoffman, A. S. *J. contr. Rel.* **6**, 297-305 (1987).
- Brannon-Peppas, L. & Peppas, N. A. *J. contr. Rel.* **8**, 267-274 (1989).
- Irie, M. *Adv. Polym. Sci.* **94**, 28-67 (1990).
- Kabra, B. & Gehrke, S. H. *Polymer Commun.* **32**, 322-323 (1991).
- Kwon, I. C., Bae, Y. H. & Kim, S. W. *Nature* **354**, 291-293 (1991).
- Kokufata, E., Zhang, Y. Q. & Tanaka, T. *J. Biomater. Sci. Polym. Ed.* **6**, 35-40 (1994).
- Okuzaki, H. & Osada, Y. *J. Biomater. Sci. Polym. Ed.* **5**, 485-496 (1994).
- Chen, G. H. & Hoffman, A. S. *Nature* **373**, 49-52 (1995).
- Yoshida, R. et al. *Nature* **374**, 240-242 (1995).
- Monji, N. & Hoffman, A. S. *Appl. Biochem. Biotech.* **14**, 107-120 (1987).
- Chen, J. P. & Hoffman, A. S. *Biomaterials* **11**, 625-630, 631-634 (1990).
- Monji, N. et al. *Biochem. Biophys. Res. Commun.* **172**, 652-660 (1990).
- Yang, H. J., Cole, C. A., Monji, N. & Hoffman, A. S. *Polymer Sci.* **A28**, 219-226 (1990).
- Chen, G. H. & Hoffman, A. S. *ACS Polym. Preprints* **33**, 468-469 (1992).
- Chen, G. H. & Hoffman, A. S. *Bioconj. Chem.* **4**, 509-514 (1993).
- Takei, Y. G. et al. *Bioconj. Chem.* **4**, 42-46 (1993).
- Monji, N., Cole, C. A. & Hoffman, A. S. *J. Biomater. Sci. Polym. Ed.* **5**, 407-420 (1994).
- Chen, G. H. & Hoffman, A. S. *J. Biomater. Sci. Polym. Ed.* **5**, 371-382 (1994).
- Chen, G. H. & Hoffman, A. S. *Macromolec. Chem. Phys.* **196**, 1251-1259 (1995).
- Chilkoti, A., Chen, G., Stayton, P. S. & Hoffman, A. S. *Bioconj. Chem.* **5**, 504-507 (1994).
- Stayton, P. S., Fisher, M. & Sligar, S. G. *J. biol. Chem.* **267**, 13544-13548 (1988).
- Weber, P. C., Ohlendorf, D. H., Wendoloski, J. J. & Salemme, F. R. *Science* **243**, 85-88 (1989).
- Burns, J. A., Butler, J. C., Moran, J. & Whitesides, G. M. *J. org. Chem.* **56**, 2648-2650 (1991).
- Chilkoti, A., Tan, P. & Stayton, P. S. *Proc. natn. Acad. Sci. U.S.A.* **92**, 1754-1758 (1995).
- Kraulis, P. J. *appl. Crystallogr.* **24**, 946-950 (1991).

ACKNOWLEDGEMENTS. This work was supported by The Whitaker Foundation (P.S.S.) and the US NSF (A.S.H.).

## A plausibly prebiotic synthesis of phosphonic acids

R. M. de Graaf, J. Visscher & Alan W. Schwartz\*

Evolutionary Biology Research Group, Faculty of Science, University of Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands

THE insolubility of calcium phosphate in water is a significant stumbling block in the chemistry required for the origin of life<sup>1</sup>. The discovery of alkyl phosphonic acids in the Murchison meteorite<sup>2</sup> suggests the possibility of delivery of these water-soluble, phosphorus-containing molecules by meteorites or comets to the early Earth. This could have provided a supply of organic phosphorus for the earliest stages of chemical evolution; although probably not components of early genetic systems, phosphonic acids may have been precursors to the first nucleic acids<sup>3</sup>. Here we report the synthesis of several phosphonic acids, including the most abundant found in the Murchison meteorite, by ultraviolet irradiation of orthophosphorous acid in the presence of formaldehyde, primary alcohols, or acetone. We argue that similar reactions might explain the presence of phosphonic acids in Murchison, and could also have occurred on the prebiotic Earth.

Methyl and ethyl phosphonic acids, although present in relatively modest concentration in water extracts (15 nmol per g meteorite), correspond to 29% of the Murchison meteorites total water-soluble phosphate<sup>2</sup>. Alkyl phosphonic acids can be regarded as being related to orthophosphorous acid, via substitution of the P-H grouping by a P-C bond (Fig. 1). It has been suggested that the meteoritic phosphonic acids could have been derived from phosphorus ylides of interstellar origin as a result of processing in a meteorite parent body<sup>2</sup>. An additional source

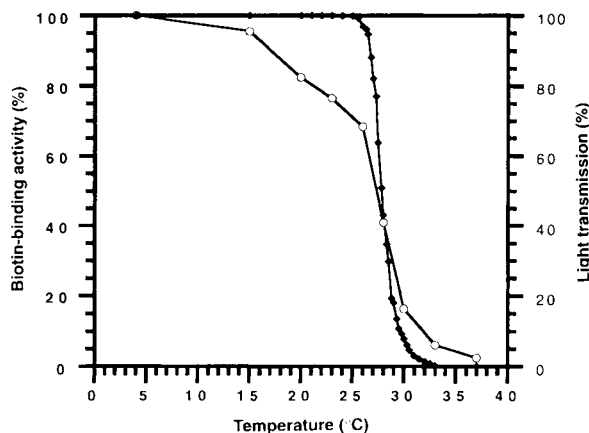


FIG. 5 Relationship between the temperature dependence of the biotin-blocking activity of the poly(NIPAAm)-streptavidin conjugate, and the temperature-stimulated cloud-point behaviour of free poly(NIPAAm). The percentage change in the biotin-binding capacity of the immobilized polymer-protein conjugate from 4 °C to 37 °C is plotted on the y-axis, assuming 100% binding (no blocking) at 4 °C and 0% binding (complete blocking) at 37 °C. The left-hand y-axis (open circles) is calculated as follows:  $[\text{H-biotin}_{(T^{\circ}\text{C})} - \text{H-biotin}_{(37^{\circ}\text{C})}] / [\text{H-biotin}_{(4^{\circ}\text{C})} - \text{H-biotin}_{(37^{\circ}\text{C})}] \times 100$ . The cloud-point behaviour is plotted as percentage light transmission of a 1% solution of the free vinylsulphone-terminated poly(NIPAAm) in 100 mM phosphate buffer (pH 7.0) from 4 °C to the endpoint at 37 °C. The right hand y-axis (filled diamonds) is calculated as follows:  $(\Delta\text{transmission}_{(T^{\circ}\text{C} - 37^{\circ}\text{C})} / \Delta\text{transmission}_{(4^{\circ}\text{C} - 37^{\circ}\text{C})}) \times 100$ .

\* To whom correspondence should be addressed.

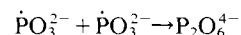


of these compounds could be via the production of  $(\dot{\text{P}}\text{O}_3)^{2-}$  radicals from phosphorous acid or phosphite. Arguments in favour of a possible prebiotic role for phosphorous acid and phosphite salts on the primitive Earth have been presented earlier<sup>1</sup>. However, these were based on the possibility of reduction of orthophosphate to phosphite in a primitive reducing atmosphere, and were criticized because the direct reduction of orthophosphoric acid is thermodynamically unfavourable<sup>4</sup>. A scenario in which phosphorous acid is produced by hydrolysis of the mineral schreibersite is much more probable. Schreibersite, a common mineral in iron meteorites, is also one of the earliest phases to be formed in a homogeneous accretion model of a cooling solar nebula<sup>5</sup>. Interaction of schreibersite with water to produce hypophosphorous<sup>1</sup> and orthophosphorous acids could plausibly occur either on a meteorite parent body, or in ablating comets or meteorites. The orthophosphorous acid thus produced could then serve as a source for photochemical production of phosphite radicals, which in turn could be used in the synthesis of phosphonic acids.

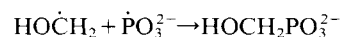
Ultraviolet irradiation was carried out with a Hanovia low-pressure mercury vapour arc tube producing light primarily at 254 nm, but with a significant component at 185 nm. We first studied reactions involving phosphite and formaldehyde, as an example of an indisputably prebiotic reactant. The irradiation of a 1 mM solution of  $\text{Na}_2\text{HPO}_3$  containing 1 mM formaldehyde ( $\text{H}_2\text{CO}$ ) in a 5.0-cm quartz cell, produced a 21% conversion of phosphite to hydroxymethyl phosphonic acid (HMPA) after ~8 h (Table 1 and Figs 2, 3). About 1% of the phosphite was converted to 1-hydroxyethyl phosphonic acid (1-HEPA). Rates of consumption of formaldehyde and production of phosphate were roughly linear up to ~6 h (Fig. 3). When we carried out irradiations in solutions that were 10 mM each in formaldehyde and sodium phosphite, the maximum yield of 1-HEPA increased to ~3%. In separate experiments in which the lamp was immersed directly in stirred solutions, up to 30% yields of HMPA were obtained within one hour. Identification of HMPA and 1-HEPA was based on combined gas chromatography-mass spectrometry of the products (GC-MS), the characteristics of which were identical to those of synthetically prepared standards. These two products were also formed by irradiation of phosphite in the presence of 1 mM methanol and ethanol, respectively.

The photolysis of aqueous formaldehyde under the conditions described here probably involves secondary reactions of hydrated formaldehyde with the primary products of the decomposition of water at 185 nm (ref. 6). The products we have identified suggest that photoreduction products of formaldehyde are trapped by recombination with phosphite radicals. The production of  $\text{PO}_3^{2-}$  radicals from sodium phosphite has previously been shown to result in the formation of hypophosphate when

phosphite is irradiated in the absence of other reactive compounds<sup>8,9</sup>.



Hydroxymethyl phosphonic acid could be produced from recombination of phosphite radicals with the hydroxymethyl radical<sup>10</sup>:



Similarly, the formation of 1-hydroxyethyl phosphonic acid requires a reduced two-carbon species, for which precedents are also available<sup>10,11</sup>. When we irradiated 1 mM aqueous solutions of primary alcohols in the presence of 1 mM phosphite, we were able to identify hydroxymethyl and 1-hydroxyethyl phosphonic acids as the major products of methanol and ethanol, respectively. We also tentatively identified the  $\alpha$ -hydroxyalkyl phosphonic acids corresponding to *n*-propanol and *n*-butanol when these alcohols were present.

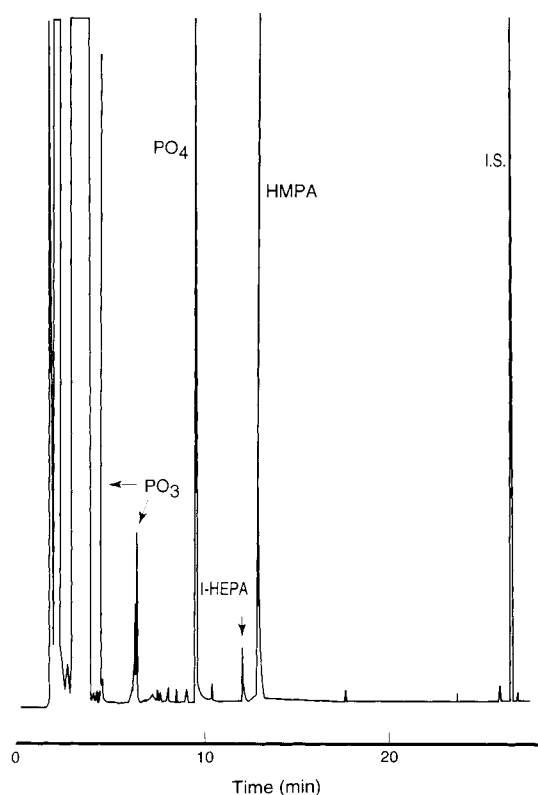


FIG. 2 Capillary gas chromatogram showing the trimethylsilyl derivatized products of the irradiation of 1 mM  $\text{H}_2\text{CO}$  in the presence of 1 mM  $\text{Na}_2\text{HPO}_3$  for 8 h (5.0 cm cell). The peaks are labelled as follows:  $\text{PO}_4$ , the di(trimethylsilyl) and tri(trimethylsilyl) derivatives of phosphorous acid;  $\text{PO}_3$ , tri(trimethylsilyl) orthophosphoric acid; HMPA, tri(trimethylsilyl) derivative of hydroxymethyl phosphonic acid; 1-HEPA, tri(trimethylsilyl) derivative of 1-hydroxyethyl phosphonic acid; I.S., internal standard (phenanthrene). Before analysis, reaction mixtures were treated with  $\text{NaBH}_4$  and silylated: to a sample of the reaction product (5.0 ml of 1 mM solutions) was added 20 mg  $\text{NaBH}_4$  and the solution was stirred at room temperature for 1 h to destroy residual formaldehyde. The solution was then passed through a column of Dowex 50W-X8 ( $\text{H}^+$ , 50–100 mesh,  $10 \times 0.8$  cm), evaporated under vacuum, and co-evaporated three times with addition of methanol (2 ml) to remove boric acid. To dried aliquots containing 20% of the product, 20  $\mu\text{l}$  dimethylformamide (containing 0.01 M phenanthrene) and 40  $\mu\text{l}$  *N*-methyl-*N*-trimethylsilyltrifluoroacetamide were added. After 15 min at room temperature, 1  $\mu\text{l}$  was injected for gas chromatographic analysis. Capillary gas chromatography was performed on a 30-m DB-17 column (0.25 mm), using a temperature programme of 80–230  $^\circ\text{C}$  at 5  $^\circ\text{C}$  per min with He as carrier gas. Peak detection was by flame ionization.

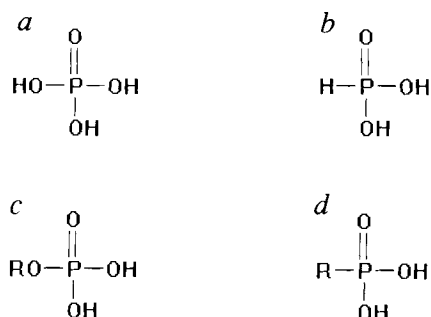


FIG. 1 Structures of orthophosphoric acid (a), orthophosphorous acid (b), an orthophosphate ester (c) and a phosphonic acid (d).

TABLE 1 Products from irradiation of 1 mM H<sub>2</sub>CO and 1 mM Na<sub>2</sub>HPO<sub>3</sub>

|                                | Chemical yield* (%) | Quantum yield† |
|--------------------------------|---------------------|----------------|
| Orthophosphate                 | 43                  | 0.20           |
| Hydroxymethyl phosphonic acid  | 21                  | 0.10           |
| 1-hydroxyethyl phosphonic acid | 1.3                 | 0.007          |

\* The chemical yields are based on phosphorus and were determined after 8 h irradiation. Orthophosphate analyses were performed by the method of Chen *et al.*<sup>15</sup>. HMPA and 1-HEPA were determined by capillary gas chromatography as described in Fig. 2.

† For determination of overall quantum yields, irradiations were conducted in a 5.0-cm-long quartz cell contained in a thermostated holder maintained at 25 °C, which was permanently mounted against the surface of the lamp (3 mm air gap). The entrance window of the holder was ~2 cm<sup>2</sup>. The radiation flux at 185 nm in the cell was determined from the production of formaldehyde from 2 M ethylene glycol. Using a quantum yield of 0.194 (ref. 16), we measured the flux as  $4.3 \times 10^{-9}$  einstein min<sup>-1</sup> ml<sup>-1</sup>, based on the slope of the initial portion of the curve. Stock formaldehyde solutions were prepared by distillation from paraformaldehyde and were standardized by titration with sodium sulphite<sup>17</sup>. Solutions were ultrasonically degassed before transfer to the irradiation vessel, were bubbled for 30 min with argon, and were magnetically stirred during the irradiation.

The production of hydroxymethyl phosphonic acid and 1-hydroxyethyl phosphonic acid from 1 mM formaldehyde and sodium phosphite would seem to be a reasonable prebiotic process. To model the possible production of methyl phosphonic acid, which is the phosphonic acid detected in highest concentration in the Murchison meteorite, we irradiated 1 mM phosphite in the presence of 1 mM acetone, which is a photochemical source of methyl radicals<sup>12</sup>. Figure 4 illustrates the identification of methyl phosphonic acid (confirmed by CG-MS comparison with authentic methyl phosphonic acid, which was produced in 17% yield after 8 h of irradiation (5.0-cm cell).

Phosphonic acids are interesting prebiotic compounds from several points of view. Compounds containing the C-P bond are extremely stable and survive conditions which would result in extensive hydrolysis of esters of phosphoric acid<sup>7</sup>. Methylene

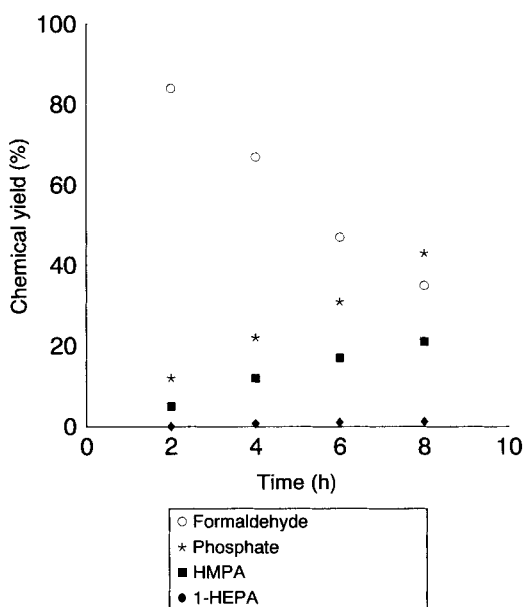


FIG. 3 Chemical yields as a function of time in the irradiation of 1 mM H<sub>2</sub>CO in the presence of 1 mM Na<sub>2</sub>HPO<sub>3</sub>. Analysis for HMPA and 1-HEPA as in Fig. 2. Formaldehyde was analysed as described previously<sup>18</sup>.

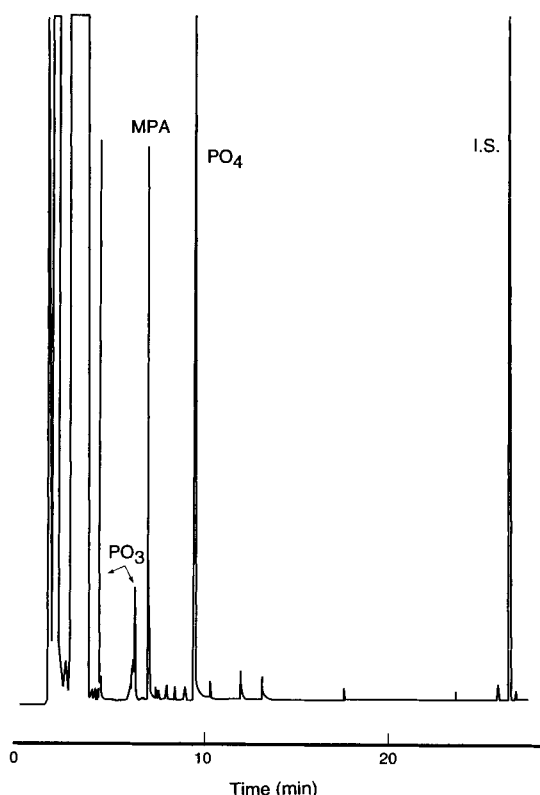


FIG. 4 Capillary gas chromatogram showing the trimethylsilyl derivatized products of the irradiation of 1 mM Na<sub>2</sub>HPO<sub>3</sub> in the presence of 1 mM acetone for 8 h (5.0-cm cell). The peak labelled MPA corresponds to the di(trimethylsilyl) derivative of methyl phosphonic acid (other abbreviations and conditions as in Fig. 2).

phosphonic acids are isosteric with orthophosphate esters, and have been synthesized as analogues of nucleotides<sup>13</sup>. Here we have used ultraviolet irradiation to model the process, but it also seems reasonable that the synthesis of phosphonic acids via the production of phosphite radicals may have occurred in the near-surface region of the meteorite parent body, where both liquid water and a source of ionizing radiation would have been present<sup>14</sup>. The presence of alkyl phosphonic acids in meteorites, as well as their ease of synthesis in an early hydrosphere of at least partially exogenous origin, would have obviated the problems inherent in the initial solubilization of apatite as a prerequisite for the synthesis of phosphate esters (preliminary experiments with ethane phosphonic acid indicate that the solubility of its calcium salt at pH 8 is greater than 0.01 M, which is the present oceanic calcium concentration). From the point of view of origin-of-life studies, there is considerable interest in the possibility that the first RNA molecules may have been preceded in evolution by the emergence of a chemically more 'robust' system, and a number of different alternative backbones for potential early genetic systems have already been examined<sup>13</sup>. A primitive precursor of RNA in which phosphates were replaced by phosphonic acids would therefore seem to be an interesting possibility, suggesting that phosphonic acids might have played a role in chemical evolution, analogous to, but before, the emergence of the first biophosphates. □

Received 4 September; accepted 3 November 1995.

1. Gulick, A. *Am. Scient.* **43**, 479-489 (1955).
2. Cooper, G. W., Onwo, W. M. & Cronin, J. R. *Geochim. cosmochim. Acta* **56**, 4109-4115 (1992).
3. Joyce, G. F. *Nature* **338**, 217-224 (1989).
4. Miller, S. L. & Urey, H. C. *Science* **130**, 245-251 (1959).
5. Fegley, B. & Lewis, J. S. *Icarus* **41**, 439-455 (1980).

6. Park, H.-R. & Getoff, N. J. *Photochem. Photobiol.* **43**, 155–163 (1988).
7. Freedman, L. D. & Doak, G. O. *Chem. Rev.* **57**, 479–573 (1957).
8. Schwartz, A. W. & van der Veen, M. *Inorg. nucl. Chem. Lett.* **9**, 39–41 (1973).
9. Matsuura, N., Yoshimura, M., Takizawa, M. & Sasaki, Y. *Bull. chem. Soc. Japan* **44**, 1027–1030 (1971).
10. Seifert, K.-G. & Bargon, J. *Angew. Chem. int. Edn engl.* **12**, 763–764 (1973).
11. Schwartz, A. W. & de Graaf, R. M. *Tetrahedron Lett.* **34**, 2201–2202 (1993).
12. Calvert, J. C. & Pitts, J. N. *393 Photochemistry* (Wiley, New York, 1966).
13. Engel, R. *Chem. Rev.* **77**, 349–367 (1977).
14. Bunch, T. E. & Chang, S. *Geochim. cosmochim. Acta* **44**, 1543–1577 (1980).
15. Chen, P. S., Toribara, T. Y. & Warner, H. *Analyt. Chem.* **28**, 1756–1758 (1956).
16. Van der Linde, H. J. & v. Sonntag, C. *Photochem. Photobiol.* **13**, 147–155 (1971).
17. Walker, J. F. *Formaldehyde* (Krieger, Huntington, NY, 1964).
18. Schwartz, A. W. & de Graaf, R. M. *J. molec. Evol.* **36**, 101–106 (1993).

ACKNOWLEDGEMENTS. This work was partially supported by NASA.

## Large-scale motion between Pacific and Atlantic hotspots

John A. Tarduno\* & Jeff Gee†

\* Department of Earth and Environmental Sciences,  
University of Rochester, Rochester, New York 14627, USA

† Scripps Institution of Oceanography, La Jolla, California 93093, USA

STUDIES of true polar wander (TPW), the rotation of the solid Earth with respect to the spin axis<sup>1</sup>, have suggested that there has been 10–15° of relative motion over the past 130 Myr (refs 2–4). In such studies, the orientation of the spin axis is recovered from continental palaeomagnetic poles (corrected for relative plate motions), and compared with a deep-mantle reference frame defined by hotspot locations. But deducing relative plate motions becomes increasingly difficult for older (Mesozoic) time periods, hindering tests of TPW on timescales comparable to those of large-scale mantle convection; moreover, the assumption of hotspot fixity is controversial<sup>5,6</sup>. We examine here a more direct approach<sup>7,8</sup>, using palaeolatitudes derived from Pacific guyots. Contrary to predictions from TPW models, these data suggest only minor latitudinal shifts of Pacific hotspots during the Cretaceous period. Instead of TPW, relative motion between the Atlantic and Pacific hotspot groups<sup>9</sup> is required at a velocity of approximately 30 mm yr<sup>-1</sup>, more than 50% larger than previously proposed<sup>5</sup>.

The Pacific Ocean basin, with its striking examples of hotspot-related topography, offers many opportunities for TPW tests. In principle, the Pacific apparent polar wander path can be compared to predictions of hotspot models<sup>10</sup>. But these tests are difficult in practice because only oceanic datasets are available from the Pacific plate. Apparent polar wander paths that include some deep-sea sediment data<sup>11</sup> are unsuitable because compaction can cause an inclination bias<sup>12,13</sup>, mimicking TPW. Paths derived from seamount palaeopoles and magnetic anomaly skewness also suffer systematic errors<sup>14,15</sup>.

These problems can be avoided if sufficient palaeomagnetic data from well dated basalts are available. For example, palaeomagnetic analyses of 65-Myr-old flows drilled in Suiko seamount (DSDP Site 433) (Fig. 1a), part of the Hawaiian–Emperor chain emanating from the active Hawaiian hotspot (19° N), indicate<sup>8</sup> a palaeolatitude of 27° N. Southward motion of 8° is thus implied<sup>16</sup>. Although close to the limit of palaeomagnetic detection, coeval northward motion is suggested from analyses of basaltic features linked to Reunion hotspot in the Indian Ocean<sup>17</sup>, consistent with TPW about an equatorial rotation pole at 110° E (ref. 18). Further progress towards testing this and older TPW rotations has been hampered by the lack of deep-penetration basement holes on seamounts which sample sufficient time to average palaeosecular variation. Ocean Drilling Program Legs 143, 144 and 145 recently drilled seamounts in the central, western and northern Pacific Ocean, affording the opportunity to conduct new tests of Cretaceous TPW.

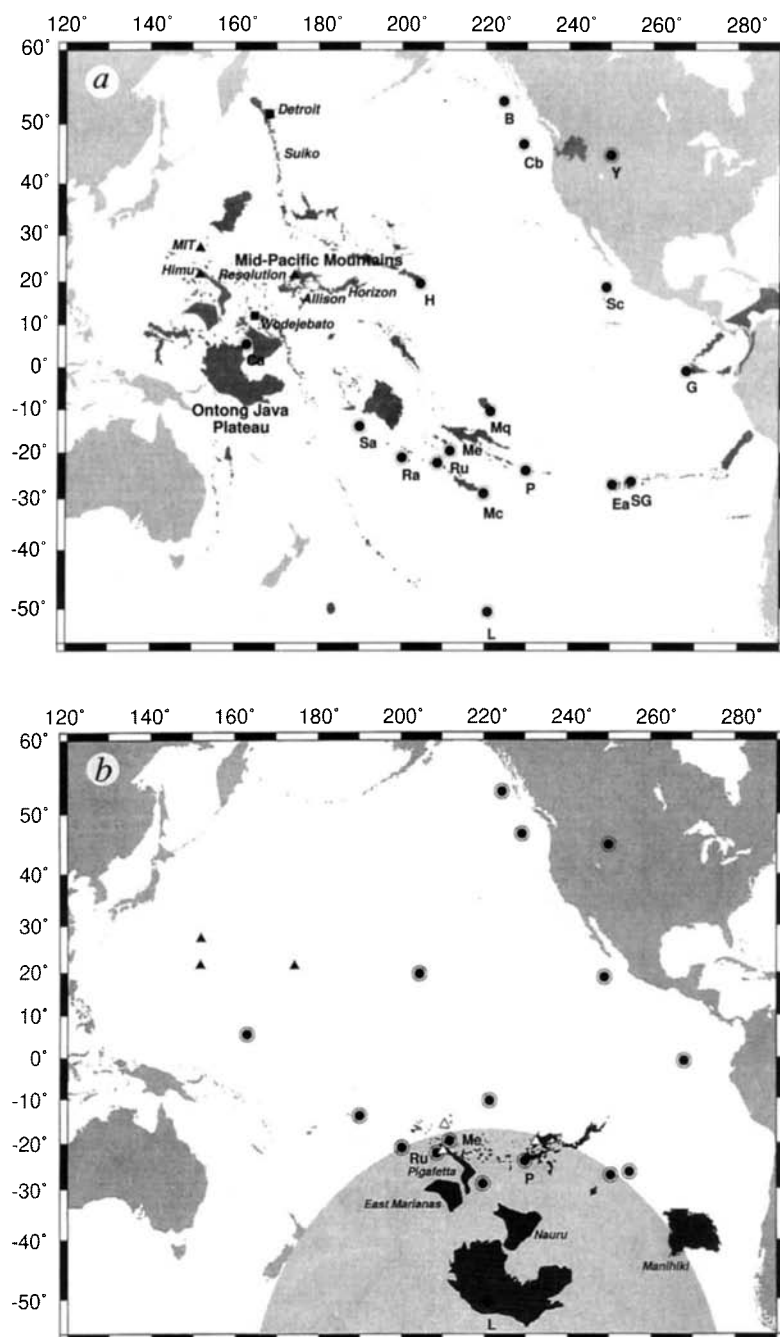
On Resolution guyot (Site 866A) in the Mid-Pacific Mountains (Fig. 1a), 126 m of altered basalt separated by weathering horizons was drilled<sup>19</sup>. The weathering horizons have magnetizations that match those of the overlying flows, and a reversal is recorded at the base of the recovered section (M7r→M8n); these observations indicate that a primary magnetization is preserved<sup>20</sup>. Magnetostratigraphy<sup>21</sup> suggests that the basalt spans ~200 kyr; the estimated directional angular dispersion ( $S = 10^\circ$ ) is within the range expected for the Early Cretaceous<sup>22</sup>. <sup>39</sup>Ar/<sup>40</sup>Ar analyses yield two age groups (123 and 128 Myr)<sup>23</sup>. Because the age data do not correlate with either the magnetic polarity or inclination, the younger age could reflect alteration. The resulting palaeolatitude is  $14.2 \pm 3.2^\circ$  S.

On MIT guyot (Site 878A) in the western Pacific close to the Wake Group of guyots (Fig. 1a), 185 m of basalt flows were drilled<sup>24</sup>. The upper flow units are of reversed polarity (M1) whereas those below are of normal polarity (M2)<sup>25</sup>. The two polarities are separated by transitional directions. These observations support both a primary magnetization and a time-sequence of sufficient duration to average palaeosecular variation. This conclusion is also supported by the estimated angular dispersion ( $S = 10^\circ$ ). An <sup>39</sup>Ar/<sup>40</sup>Ar age of 124 Myr has been obtained from the sequence<sup>26</sup>. Overall, the data suggest a palaeolatitude of  $11.5 \pm 2.3^\circ$  S.

Comparison of these direct palaeolatitude determinations with palaeolatitudes predicted from plate reconstructions provides a first-order test of hotspot fixity and TPW. Using the hotspot reconstruction poles of Duncan and Clague<sup>27</sup>, Resolution guyot seems to have formed 850 km northeast of Pitcairn hotspot at 128 Myr, whereas MIT formed 400 km NNE of Mehitia hotspot at 124 Myr (Fig. 1b). This reconstruction also implies formation of Himu seamount (<sup>39</sup>Ar/<sup>40</sup>Ar age of 120 Myr) slightly northeast of the isotopically similar Rurutu hotspot<sup>28</sup>. The geochemical similarity of these features suggests that a link to extant hotspots may not be unreasonable for some of the western Pacific seamounts. We consider below modifications of the hotspot reconstructions that link guyots to extant hotspots, although it is possible that these Early Cretaceous guyots formed far from hotspots or were formed by hotspots that have waned and have no Cenozoic trace.

We first examine linking Resolution guyot directly to Pitcairn hotspot in the Early Cretaceous (128 Myr) (Fig. 1c). When the new palaeolatitude estimate for Resolution guyot is compared with the latitude of the Pitcairn hotspot, a 10° S offset is seen. But in this reconstruction there is a longitudinal discrepancy between Mehitia hotspot and MIT guyot, and a gap in both latitude and longitude between Rurutu hotspot and Himu seamount (Fig. 1c). If MIT is matched exactly with Mehitia hotspot, Himu seamount falls within 500 km of Rurutu seamount, but Resolution guyot is far to the east of Pitcairn hotspot, and far from any other hotspot. The former reconstruction is more consistent with Cenozoic volcanism. On the Hawaiian chain, recent volcanism has been observed on edifices (Kauai and Niihau) that are 5 Myr older than the locus of present-day volcanism<sup>29</sup>. An eruptive duration of 5 Myr could explain the offsets in our first reconstruction, but fails to explain the mismatches in the alternative reconstruction because Resolution would not yet have passed over Pitcairn hotspot.

These mismatches could record relative motion between the three hotspots; alternatively, they might reflect unrecognized crustal extension between the guyots and seamounts after their formation related to formation of the Ontong Java Plateau (Fig. 1b). Geochemical analyses support the creation of ephemeral ridge segments in other Jurassic basins during the Aptian volcanic episode<sup>30</sup>. If correct, MIT may have formed closer to Mehitia hotspot than suggested by the reconstruction. When MIT is reconstructed to Mehitia hotspot, an offset of 8° S is observed between the new palaeolatitude estimate from MIT guyot and the location of Mehitia hotspot, similar to that observed from Resolution guyot.



Two Late Cretaceous data sets have also recently become available. During Leg 144, five sites were drilled on Wodejebato guyot in the southwestern Pacific (Sites 873–877) (Fig. 1a)<sup>31</sup>. Although the deepest penetration at any site was only 50 m, palaeomagnetic analyses of all the recovered basalt yields seven reversed polarity units (33R)<sup>25</sup> which together suggest a palaeolatitude of  $8.4 \pm 4.6^\circ$  S. The estimated angular dispersion ( $S = 12^\circ$ ) is slightly higher than expected for rocks of this age<sup>22</sup> and may reflect complications caused by unrecognized tilting or differences in magnetization age. Hotspot reconstructions suggest that Wodejebato guyot was  $23^\circ$  S  $\sim 80$  Myr ago. Although nearest to Rarotonga hotspot<sup>27</sup> at this time, the sampled basalts formed on a much older edifice<sup>19</sup>. An offset of  $14.6^\circ$  S is observed between the new palaeolatitude estimate and the hotspot model latitude.

Detroit seamount is part of the Hawaiian–Emperor chain near the Aleutian–Kuril trench (Fig. 1a). During Leg 145, a single hole penetrated 87 m of basalt which contained 13 units, of which 10 were massive flows<sup>32</sup>. Shipboard palaeomagnetic data

from the site<sup>32</sup> yielded 11 inclination units of reversed polarity which together define a mean latitude of  $33.3^\circ \pm 8.6^\circ$  N. The estimated angular dispersion is  $\sim S = 17^\circ$ . The palaeolatitude estimate differs by  $14^\circ$  S from the latitude of the Hawaiian hotspot.  $^{39}\text{Ar}/^{40}\text{Ar}$  radiometric data suggest an age of 81 Myr for the sampled volcanics<sup>33</sup>. More complete palaeomagnetic analyses are needed for Detroit seamount. Nevertheless, the available results agree with those from Wodejebato guyot.

The offsets between the spin axis and hotspots defined by the new Early Cretaceous Pacific data are only slightly different from those known from Suiko seamount (65 Myr), and therefore disagree sharply with TPW motion predicted from global continental data<sup>4</sup> (Fig. 2). Although the data from Detroit seamount and Wodejebato hint that some northward drift of the Pacific hotspots might have occurred at or near the end of the Cretaceous normal-polarity superchron, the offset between the spin axis and Pacific hotspots is nearly constant over the interval from 130–95 Myr. Because we have no reason to doubt the continental data, the large discrepancy suggests relative motion between the



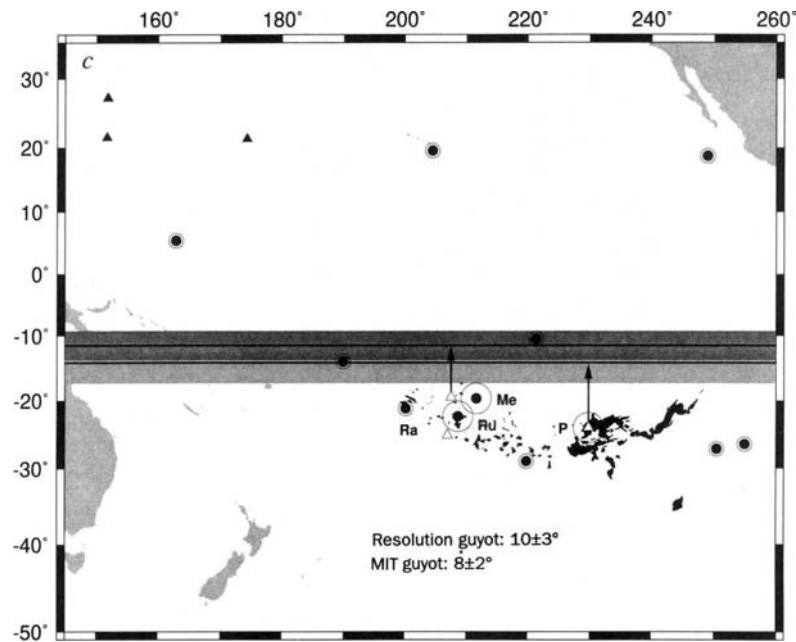
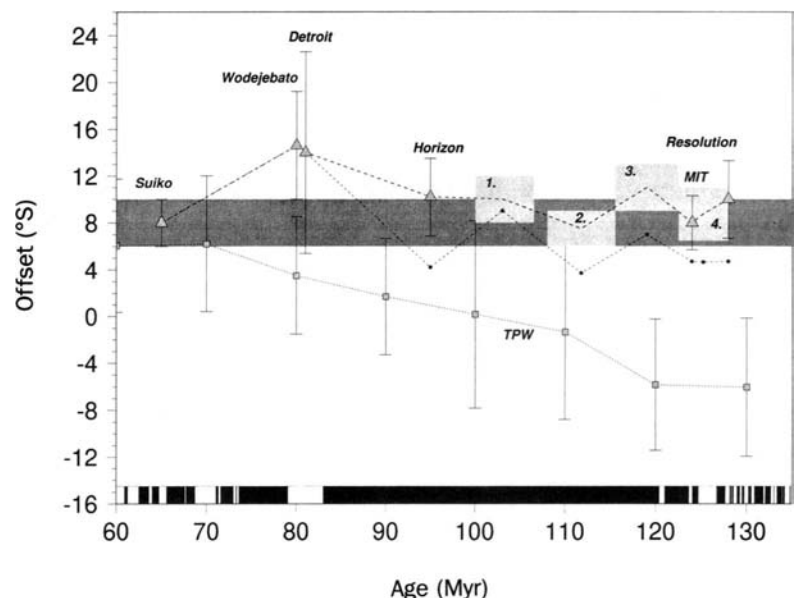


FIG. 1 a, Pacific basin showing locations of hotspots<sup>27,28</sup> and large igneous provinces<sup>37</sup>. B, Bowie; Ca, Caroline; Cb, Cobb; Ea, Easter; G, Galapagos; H, Hawaii; L, Louisville; Me, Mehitia; Mc, MacDonald; Mq, Marquesas; P, Pitcairn; Ra, Rarotonga; Ru, Rurutu; Sa, Samoa; Sc, Socorro; SG, Sala y Gomez; Y, Yellowstone. Also shown are seamounts and guyots discussed: Early Cretaceous (filled triangles), Late Cretaceous (filled squares) and Cretaceous/Tertiary boundary (open square). b, Reconstruction (Mercator projection) based on fixed Pacific hotspots<sup>27</sup>; continents shown in present-day positions. Filled triangles, positions of MIT and Resolution guyot and Himu seamount (see Fig. 1), open triangles, position at 120 Myr. At 120 Myr Ontong-Java Plateau and Himu seamount reconstruct to the Louisville and Rurutu hotspots<sup>28</sup>, respectively. Volcanism in the Nauru basin, Pigafetta basin, East Marianas basin and Manihiki Plateau are related to formation of the Louisville hotspot<sup>35</sup>. Areas to the southwest of Ontong Java Plateau

have been subducted. Grey semicircle marks the inferred location of a mantle plume head<sup>35</sup>. Detroit seamount and Wadejebato guyot form after the time shown in this reconstruction. The former is assumed to have formed at the Hawaiian hotspot at 81 Myr while volcanics sampled at the latter erupted as the guyot moved near Rarotonga hotspot at ~80 Myr. c, Mid-Pacific Mountains (Resolution guyot) reconstructed to Pitcairn hotspot at 128–120 Myr with a total reconstruction pole at  $47.1^\circ$  N,  $277.6^\circ$  E and an anticlockwise rotation angle of  $71.2^\circ$ ; continents shown in present-day positions. Solid lines show palaeomagnetically determined palaeolatitudes from MIT and Resolution guyots; shaded bars represent 95% confidence regions. Arrows represent offsets between hotspot latitudes and palaeomagnetic determinations. Circles around Mehitia, Rurutu and Pitcairn hotspots have a diameter of 500 km.

FIG. 2 Offsets between hotspot predicted latitudes and palaeomagnetic latitudes for the Pacific. Values for Resolution<sup>20</sup>, MIT<sup>25</sup>, Wadejebato<sup>25</sup>, Detroit<sup>32</sup> and Suiko<sup>8</sup> are from direct palaeomagnetic measurement; Horizon is calculated from a synthetic Pacific apparent polar wander path<sup>20</sup>. Large shaded boxes represent basal sediments from Mid-Pacific Mountain sites<sup>20</sup>: 1, Allison guyot, Site 865; 2, 3, Resolution guyot, sedimentary apron, Site 463; 4, Resolution guyot. Triangles represent guyots linked to individual Cenozoic hotspots (see text), filled circles use hotspot model latitudes<sup>27</sup>. Offsets recorded by Cretaceous guyots using either approach are similar to those recorded by Suiko seamount (shaded bar) but differ from offset of the Hawaiian hotspot (boxes) predicted from true polar wander model of Besse and Courtillot<sup>4</sup>. Error bars are 95% confidence intervals. Geomagnetic timescale after Gradstein et al.<sup>39</sup>.



130-90 Myr

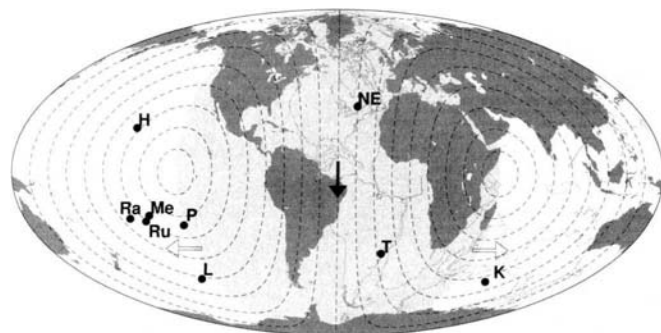


FIG. 3 Potential mantle-roll rotation (dashed lines) satisfying data from the New England hotspot track (NE) and trace of the Tristan da Cunha hotspot (T), while minimizing latitudinal motion of Pitcairn hotspot (P). Continents are shown in present-day position (solid) and position at 130 Myr (dashed), assuming no motion between Atlantic and Indian ocean hotspots<sup>38</sup>. Shaded area shows hemisphere experiencing southward motion. The southward motion of NE and T (solid arrow) predicts westward motion of Pacific hotspots and eastward motion of Indian ocean hotspots (open arrows), neither of which is observed. Instead, our preferred interpretation is that the Atlantic hotspots moved southwards while the Pacific hotspots remained stationary with respect to the spin axis.

Pacific hotspots and those in the Atlantic and Indian oceans<sup>34</sup> used to calculate TPW<sup>4</sup>. Note that the TPW predictions, which have error bars that account for errors in plate rotations and hotspot locations, are statistically distinct from the new Pacific data for times older than 110 Myr (Fig. 2).

Other lines of evidence further support a nearly constant latitudinal offset between Early Cretaceous Pacific hotspots and the spin axis (Fig. 2). The standstill of the Pacific apparent polar wander paths during the Early to mid-Cretaceous, when compared to the hotspot track defined by the Mid-Pacific Mountains<sup>20</sup>, suggests that the latitudinal offset defined by MIT and Resolution guyots should not have changed during the mid-Cretaceous. Palaeomagnetic data from basal shallow-water sediments on Mid-Pacific guyots that are thought to have escaped the effects of compaction<sup>21</sup> support this conclusion. Perhaps the most stringent constraint is provided by the data from Resolution guyot. Palaeomagnetic data from volcanic basement, basal sediments and the flanking sedimentary apron all yield statistically indistinguishable palaeolatitudes spanning a period of 128–108 Myr. The slope of the hotspot spin-axis offset curve defined by these data ( $0.02 \pm 0.04^\circ \text{ N}$ ;  $1\sigma$  error) is distinct from that predicted by TPW<sup>4</sup> ( $0.21 \pm 0.08^\circ \text{ Myr}^{-1} \text{ S}$ ), even if an additional uncertainty of 500 km is added to each of the Pacific observations. Using the latter, conservative estimate, a relative motion of  $26 \pm 19 \text{ mm yr}^{-1}$  is indicated.

Another test of this relative motion is available from the comparison of the new Pacific data to those from the Atlantic. In contrast to the Pacific data, Van Fossen and Kent<sup>9</sup> report a large  $12^\circ$  (S) difference between palaeolatitudes and the present-day New England hotspot latitude, which accumulated entirely during the mid-Cretaceous (90–124 Myr). Because the North American apparent polar wander path is stationary during the

mid-Cretaceous, the motion of the New England hotspot, if representing the entire mantle, is more properly classified as ‘mantle roll’<sup>7</sup> rather than TPW *sensu stricto* because the lithosphere is uncoupled from the underlying rotating mantle.

Because the Pacific hotspots show little or no mid-Cretaceous latitudinal motion, a mantle-roll rotation pole, which is constrained to lie on the equator by physical considerations<sup>7</sup> (shifts in the Earth’s principal moments of inertia), must lie in the Pacific basin. For illustration, we consider an equatorial pole of rotation at the longitude of Pitcairn hotspot (Fig. 3), that also satisfies the southward motion of the Atlantic hotspots. However, this rotation pole also predicts that Cretaceous volcanic chains  $\sim 6^\circ$  long should form on the Pacific plate and that these chains should become progressively younger to the west. Instead, chains such as the Mid-Pacific Mountains show an eastward younging from 128 Myr (Resolution guyot) to 95 Myr (Horizon guyot). Shifts of the rotation pole to the east or west introduce a latitudinal component of motion not seen in the new palaeomagnetic data to this longitudinal age discrepancy. In addition,  $\sim 13^\circ$  of eastward motion of the Kerguelen hotspot is also predicted by the rotation, which is incompatible with hotspot tracks in the Indian Ocean<sup>18</sup>.

Instead of TPW or mantle-roll, the palaeomagnetic data support the suggestion of large-scale relative motion between the Pacific and Atlantic hotspot groups<sup>9</sup>. The rate of motion is  $34 \pm 9 \text{ mm yr}^{-1}$ ; if we add an additional error of 500 km to each Pacific and Atlantic observation to account for gross errors in hotspot locations, the error region is larger, but the motion of  $31 \pm 20 \text{ mm yr}^{-1}$  remains statistically significant. This pulse of rapid hotspot motion coincides with large-scale volcanism in the Pacific and Indian oceans<sup>35,36</sup>; both observations may reflect a change in the mode of lower-mantle convection during the mid-Cretaceous. □

Received 10 July; accepted 16 October 1995.

- Goldreich, P. & Toomre, A. *J. geophys. Res.* **74**, 2555–2567 (1969).
- Harrison, C. G. A. & Lindh, T. *Nature* **300**, 251–252 (1982).
- Andrews, J. *J. geophys. Res.* **90**, 7737–7750 (1985).
- Besse, J. & Courtillot, V. *J. geophys. Res.* **96**, 4029–4050 (1991).
- Molnar, P. & Stock, J. *Nature* **327**, 587–591 (1987).
- Acton, G. D. & Gordon, R. G. *Science* **263**, 1246–1254 (1994).
- Hargraves, R. B. & Duncan, R. A. *Nature* **245**, 361–363 (1973).
- Kono, M. *Init. Rep. DSDP Leg 55*, 737–752 (1980).
- Van Fossen, M. C. & Kent, D. V. *J. geophys. Res.* **97**, 19651–19661 (1992).
- Gordon, R. G. *Geophys. Res. Lett.* **10**, 709–712 (1983).
- Sager, W. W. & Bleil, U. *Nature* **326**, 488–490 (1987).
- Tarduno, J. A. *Geophys. Res. Lett.* **17**, 101–104 (1990).
- Gordon, R. G. *J. geophys. Res.* **95**, 8397–8404 (1990).
- Gee, J., Staudigel, H. & Tauxe, L. *Nature* **342**, 170–173 (1989).
- Larson, R. L. & Sager, W. W. *Proc. ODP Sci. Res.* **129**, 471–481 (1992).
- Gordon, R. G. & Cape, C. D. *Earth planet. Sci. Lett.* **55**, 37–47 (1981).
- Vandamme, D. & Courtillot, V. *Proc. ODP Sci. Res.* **115**, 111–118 (1990).
- Duncan, R. A. & Storey, M. in *Synthesis of Results from Scientific Drilling in the Indian Ocean* (eds Duncan, R. A., Rea, D. K., Kidd, R. B., von Rad, U. & Weissel, J. K.) 91–103 (AGU Monogr. 70, Am. Geophys. Union, Washington DC, 1992).
- Winterer, E. L., Sager, W. W., Firth, J. V. & Sinton, J. M. *Proc. ODP Sci. Res.* **143**, (1993).
- Tarduno, J. A. & Sager, W. W. *Science* **269**, 956–959 (1995).
- Tarduno, J. A., Sager, W. W. & Nogi, Y. *Proc. ODP Sci. Res.* **143**, 395–398 (1995).

- McFadden, P. L., Merrill, R. T., McElhinny, M. W. & Lee, S. J. *Geophys. Res.* **96**, 3923–3933 (1991).
- Pringle, M. S. & Duncan, R. A. *Proc. ODP Sci. Res.* **143**, 277–283 (1995).
- Shipboard Scientific Party Init. Rep. *ODP Leg 144*, 331–412 (1993).
- Nakanishi, M. & Gee, J. *Proc. ODP Sci. Res.* (in the press).
- Pringle, M. S. *Proc. ODP Sci. Res.* (in the press).
- Duncan, R. A. & Clague, D. A. in *Ocean Basins and Margins Vol. 7A* (eds Nairn, A. E. M., Stehlé, F. G. & Uyeda, S.) 89–121 (Plenum, New York, 1985).
- Smith, W. H. F., Staudigel, H., Watts, A. B. & Pringle, M. S. *J. geophys. Res.* **94**, 10501–10523 (1989).
- Clague, D. A. & Dalrymple, G. B. *Contr. Miner. Petrol.* **99**, 202–218 (1988).
- Castillo, P. R., Pringle, M. S. & Carlson, R. W. *Earth planet. Sci. Lett.* **123**, 139–154 (1994).
- Shipboard Scientific Party Init. Rep. *ODP Leg 144*, 145–312 (1993).
- Shipboard Scientific Party Init. Rep. *ODP Leg 145*, 209–302 (1993).
- Keller, R. A., Fisk, M. R. & White, W. M. *Eos (Suppl.)* **75**, 723 (1994).
- Morgan, W. J. in *The Sea Vol. 7* (ed. Emiliani, C.) 443–487 (Wiley, New York, 1981).
- Tarduno, J. A. et al. *Science* **254**, 399–403 (1991).
- Larson, R. L. *Geology* **19**, 547–550 (1991).
- Coffin, M. F. & Eldholm, O. *Geology* **21**, 515–518 (1993).
- Muller, R. D., Royer, J.-Y. & Lawver, L. A. *Geology* **21**, 275–278 (1993).
- Gradstein, F. U. et al. *J. geophys. Res.* **99**, 24051–24074 (1994).

ACKNOWLEDGEMENTS. We thank P. Wessel and W. H. F. Smith for GMT software and G. Acton for review of the manuscript. This research was supported by the University of Rochester and the US National Science Foundation.

# The origins of Andean irrigation

K. S. Zimmerer

Department of Geography, University of Wisconsin, 384 Science Hall, 550 North Park Street, Madison, Wisconsin 53706-1491, USA

**BETWEEN AD 530 and the early 1500s, the people of the Andes built complex systems of irrigation, aqueducts and raised fields in the tropical mountains of present-day Bolivia, Ecuador and Peru. Their accomplishments supplied crops and the economic and symbolic bases of power for some of the world's most notable early civilizations, including the Empires of the Tiwanaku (~AD 100–1000) and Inka (~AD 1400–1532), as well as smaller chiefdoms<sup>1, 6</sup>. But there has been no direct evidence of extensive irrigation, the most common water control, before about AD 1000. Here we demonstrate, by studies of sedimentology and geomorphology near Tarata village in the Bolivian Andes, that an extensive and integrated system of floodwater and canal irrigation was in use at about AD 719 and that it functioned as early as 3,500 years before present (BP). A modern-day equivalent continued in use until 1993. The findings at Tarata give a detailed example of early irrigation that would have enabled the development of water control by Andean civilizations in highland settlements above 2,000 m.**

Analysis of the Tarata irrigation system is based on the following: (1) a description of the geographical setting, scope and recent functioning (1991) of hybrid floodwater-canal irrigation; (2) evidence of massive silt-size sediments deposited gradually as a result of this irrigation; (3) multifold alteration of fluvial geomorphology resulting from irrigation; and (4) age estimates of floodwater-canal irrigation determined by <sup>14</sup>C analysis of charcoal buried by the irrigation-derived sediment. Evidence of hybrid floodwater-canal irrigation occurs on the alluvial fan below Tarata village in the Cochabamba Department of highland Bolivia (Fig. 1). Long-term irrigation deposited a massive silt layer on the surface of the 26.8 km<sup>2</sup> Tarata fan. Silt was carried by the intermittent flow of sediment-laden runoff from the 115 km<sup>2</sup> upland of the main Calicanto and smaller Seco watersheds where rainfall averages 500 mm per year. The Calicanto and Seco rivers flowed north across the low-gradient Tarata fan (0.0081 slope)<sup>7</sup>. They unloaded coarse sediment in channel bottoms, while fine-grain silt reached fields on the fan's surface by a carefully designed network of offtakes and canals. Farmers operated floodwater-canal irrigation using the following features in 1991. First, diversion barriers of boulders, cobbles, sand and gravel were mounded as high as 2 m at oblique angles in the channels to shunt floodwater into headwaters of the main or 'mother' canals (Fig. 1); these barriers were rebuilt every few years. Second, four main canals with cobble-strewn bottoms and earthen banks diverged from the Calicanto channel and fed numerous branch canals—in total measuring more than 50 km—that carried silty irrigation water to fields over most of the Tarata fan (Fig. 1). Third, bunds, opened briefly to allow the inflow of irrigation, pooled water in small fields that were cropped with maize, alfalfa, potatoes and wheat by more than 2,000 Quechua-speaking peasant farmers in 1991 (most fields measured between 0.1 and 0.5 ha). This floodwater-canal

irrigation system was in use until 1993, when its use was ended abruptly by the completion of a 2,500,000-m<sup>3</sup> dam in the Calicanto channel upstream of Tarata. However, the lifetime of this new dam might be cut short as a result of severe erosion in the Calicanto watershed<sup>8,9</sup>.

Yearly use of the hybrid-style irrigation gradually deposited a massive silt layer on the Tarata fan. A striking predominance of fine particles throughout vertical sections of 4 m and more is exposed in the cut banks of river channels. Sections at the cut banks at P-11 (4.7 m), P-14A (4.0 m) and P-14B (6.0 m) on the Calicanto channel uniformly show that 60–80% of the massive deposits are silt (Fig. 2). This silt unit is up to 8 m thick, according to drillers' logs for several 50-m bore holes<sup>10</sup> (BC1–4 in Fig. 1). Its lateral contiguity is shown by 8 soil pits dug to 2 m (SP1–SP8) that were located laterally across the fan surface (Fig. 1). Silt-size particles predominate in all strata of the 8 pits<sup>11</sup>. The extensive contiguity—both vertical and lateral—of the fan's silt unit must have resulted from the long-term deposition of irrigation-derived sediments and the related canalization of the Calicanto channel. If sedimentation had not been controlled by irrigation, the Tarata fan, like other unirrigated outwash fans, would show a heterogeneous mixture of coarse materials in both its vertical and lateral sections owing to migration of the river channel across the fan's surface<sup>12–15</sup>.

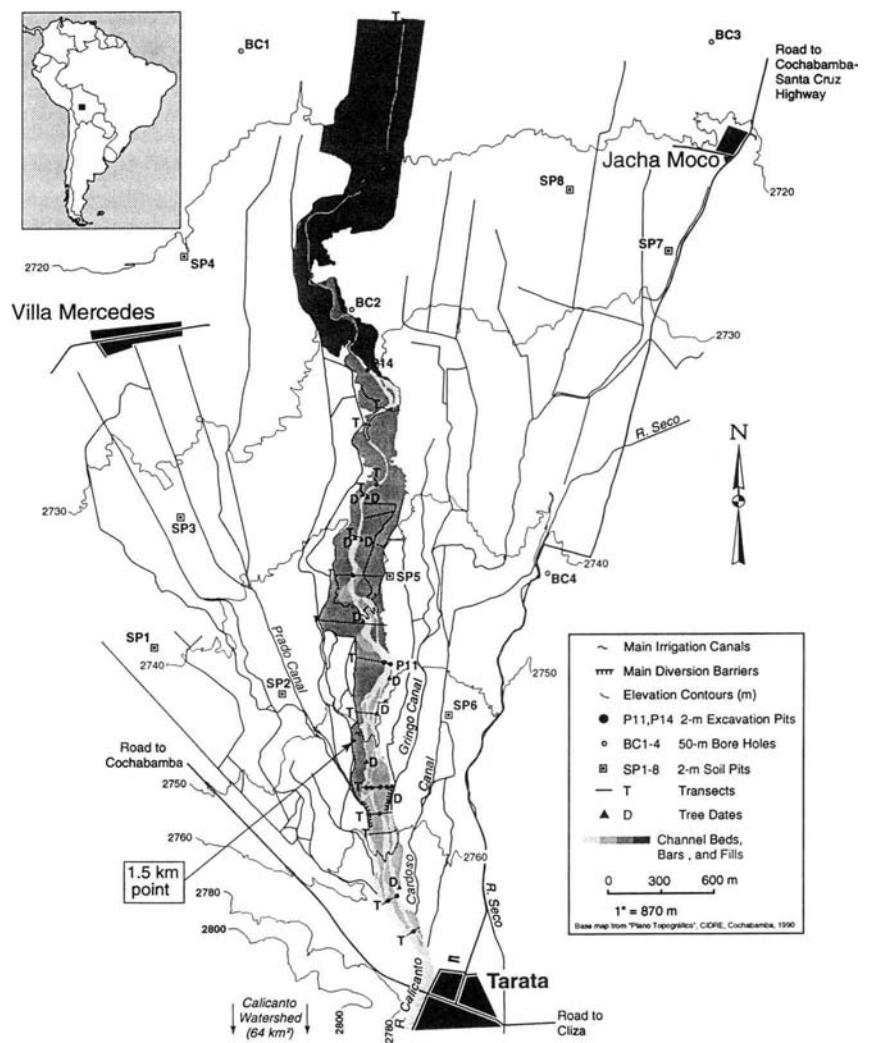


FIG. 1 An extensive system of floodwater-canal irrigation on the Tarata alluvial fan, Cochabamba, Bolivia. Andean farmers designed the main canals to radiate from the Calicanto River and to channel water and silt to fields across the fan's surface.

Further evidence of a long-term, irrigation-derived origin of the silt unit is obtained by comparing it to the fluvial geomorphology of the river channel. The Calicanto River has unloaded coarse material consisting of boulders (>5 cm), cobbles (<5 cm), and gravel as far downstream as P14 (Fig. 1). Indeed the channel surfaces at sites P11, P14-A and P-14B contain a predominance of coarse material (Fig. 2, bottom), while 15 pits excavated to 2 m at various locations in the Calicanto channel (P1–P15 marked with filled circles in Fig. 1) showed that 75% of the strata contain 20% or more coarse material (gravel, cobbles and boulders). Coarse sediment on the channel bed and the presence of coarse material under the massive silt cap of the Tarata fan affirm that coarse sediment has reached the fan throughout its history. However, the absence of channel-type deposits interbedded within this fan's silt unit indicates that long-term irrigation confined the delivery of coarse sediment to the canalized main channel, while dispersing fine sediment elsewhere across the fan surface.

The long-term effect of irrigation is also shown by the fluvial geomorphology of the Calicanto River. At 1.5 km downstream from the fan apex, the channel deposits, calculated shear stress, morphology and longitudinal profile change abruptly as a result of the energy-sapping diversion of floodwaters into main canals and the river's sudden loss of transport capacity (Fig. 1). At the 1.5-km mark the average clast diameter of channel deposits drops from 350 mm or more to less than 200 mm. Shear stress is reduced from about  $210 \text{ Nm}^{-2}$  to about  $60 \text{ Nm}^{-2}$  (ref. 16). Also at the 1.5-km mark the channel's morphology is transformed from a high-energy braided course to a single, narrow one (Fig. 1). Finally, above this point the profile of the Calicanto River exhibits an unusual inflection that is convex-up, and is the long-term consequence of abrupt bed load deposition in its upper channel where diversion barriers shunt off flow.

Age of irrigation use on the Tarata fan was determined by  $^{14}\text{C}$  analysis of charcoal fragments found buried in the irrigation-derived silt unit at three sites. At the southern P-11 site a dense concentration of diverse size charcoal fragments (sample 1) was found in the cut bank 4.7 m high between 123 and 133 cm above the channel surface (Fig. 2). Distinct segregations of fired clay surrounded the charcoal. A small, unmarked, ceramic shard was found at 138 cm. Conventional radiocarbon analysis yielded an uncalibrated date of  $1,275 \pm 65 \text{ yr BP}$  (A-7458) (see Fig. 2).

In the area of P-14, charcoal fragments were recovered at two nearby sites separated by 8 m, P-14A and P-14B (Fig. 2). The two lower strata (0–200 cm and 200–300 cm), identified on the basis of colour, structure and minor contrasts in texture, were

continuous between P-14A and P-14B. Charcoal occurred in the lowest stratum of only the P-14A section. Unsorted fragments collected between 170 and 200 cm (sample 2) at the 4.0-m-high site at P-14A were dated at  $3,665 \pm 95 \text{ yr BP}$  (A-7453) by conventional radiocarbon analysis (Fig. 2). Charcoal pieces sampled between 250 and 300 cm (sample 3) at P-14A, also variable in size, were dated using conventional techniques to  $3,880 \pm 240/235 \text{ yr BP}$  (A-7455). Charcoal was also sampled in a higher layer between 300 and 350 cm that occurred only at the 6.0-m-high site at P-14B (sample 4). Accelerator mass spectrometry yielded a date of  $4,265 \pm 60 \text{ yr BP}$  (AA-7456). Charcoal in samples 1–4 occurred with sediment identical to adjacent portions of each stratum in terms of colour, structure and texture.

The ages and stratigraphic position of  $^{14}\text{C}$ -dated samples confirm that charcoal samples were *in situ*. Dense concentrations of unsorted fragments (non-uniform size) in discrete lenses attest to their *in situ* origin. Charcoal-bearing sediment lenses extend horizontally for 0.5–2.0 m before terminating at well-defined edges. Compactness and *in situ* origin is also illustrated by the well defined vertical dimension of the lenses, which vary from 10 cm (sample 1) to 50 cm (sample 4), and their non-contiguous stacking. (Water-borne redeposition, by contrast, would sort charcoal fragments and deposit them less discretely, resulting in horizontal trailing off.)

A lack of anomalies in the charcoal-bearing sediments also indicates the charcoal's *in situ* origins. If, by contrast, the charcoal was redeposited from *ex situ* sources nearby, then the sedimentological contexts would differ from adjacent portions of each stratum. Other evidence of *in situ* origin includes the fired clay associated with sample 1 at P-11, spalled rocks mixed with sample 3 at P-14A, and distinct but similar ages at the P-14A and P-14B sites. It is most likely that the locally abundant charcoal was derived from a periodic burning of vegetation cleared from fields and the surrounding area, perhaps including the adjacent riparian zone.

Finally, overlap and partial inversion of estimated ages at the P-14A and P-14B sites (samples 2, 3 and 4) might be explained by the separation of sites and the expected reworking of field sediments with the deep-bladed Andean foot-plough. Sample 4 at P-14B, the oldest, was buried at a distance and thus separately from the other samples, probably in a different field, which would account for its higher stratigraphic position. The dates of samples 2 and 3 overlap: if error limits are considered at the 95% confidence level, sample 2 is  $3,665 \pm 190$  years and sample 3 is  $3,880 \pm 480/470$  years old. The ages of samples 2 and 3 are expected to overlap because the soils derived from silt deposits

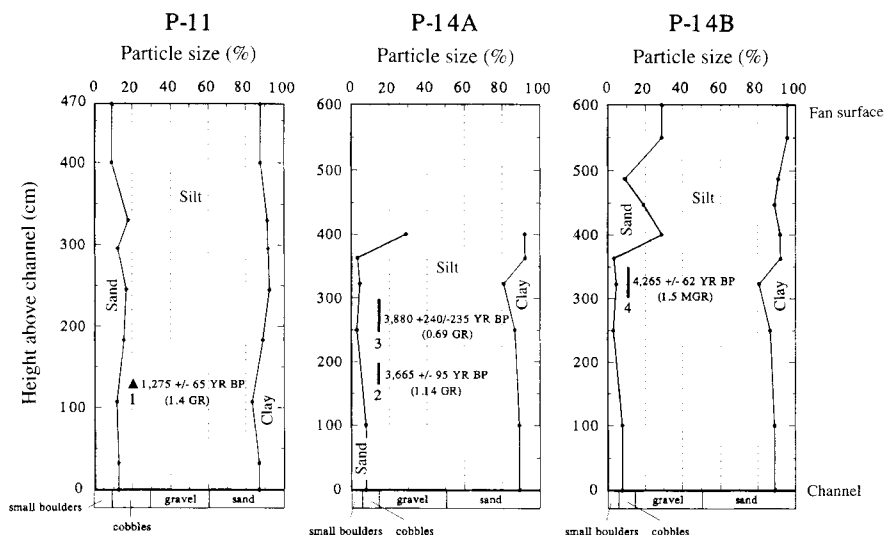


FIG. 2 Sediment texture and dated charcoal remains in cut banks of the Calicanto River channel on the Tarata alluvial fan at P-11, P-14A and P-14B. Composition of channel deposits is shown at the bottom. The amount of tested carbon is below each date. Dates are corrected for  $\delta^{13}\text{C}$ ; they are not dendro-calibrated.



on the alluvial fan are notoriously poor in organic matter, and even recently could be cultivated only with the intensive ploughing in of manure and crop stubble.

The approximate date of AD 719 thus marks a time when the Tarata waterworks were already functioning, although 3,500 yr BP is an estimate of its earliest use. Unfortunately the preceramic and early ceramic periods of the Cochabamba highlands are virtually unknown<sup>17</sup>. Remains of palaeo-canals, such as one located 50 m downstream of the Gringo Canal offtake, suggest a wealth of material for archaeological research (Fig. 1). Shard fragments found with the charcoal-bearing lens of the P-11 site indicate that by AD 719 irrigated fields were probably being used occasionally as sites of cooking and perhaps even short-term habitation, as they are in the present-day farming systems of Quechua- and Aymara-speaking peasants.

The extensive system of floodwater-canal irrigation at Tarata offers a new perspective on early Andean waterworks. It demonstrates that later civilizations such as the Tiwanaku and Inka Empires could tap a mountain-based mastery of controlling floodwater and engineering canals. A legacy of expertise would have contributed to their rapid expansion of water control and management. Hybrid arrangements, such as Tarata, were probably an important transitional stage in the evolution of Andean irrigation from rudimentary floodwater farming without canals to the full-scale canal irrigation that was typical of the later empires and chiefdoms. Finds at Tarata suggest parallels with the early history of irrigation in the coastal and low inter-Andean valleys of western South America, where farmers adopted irrigated farming by 3,500 yr BP<sup>17–20</sup>.

Evidence of early irrigation at Tarata also explains the long-term use of environments and agricultural resources in the Andean uplands. Tarata's irrigation system offers the most specific evidence so far of how various world-renowned Andean crops could be cultivated before Inka rule in the widespread and heavily populated semiarid and subhumid environments that cover intermediate elevations between 2,000 and 3,500 m (refs 21, 22). Only early irrigation, exemplified by the Tarata works, could

have made possible the expansion of farming and the culture of diverse crops in the important intermediate-elevation habitats of the tropical Andes mountains. □

Received 14 March; accepted 6 October 1995.

1. Hastorf, C. A. & Earle, T. K. in *Prehistoric Intensive Agriculture in the Tropics* Vol. 2 (ed. Farrington, I. S.) 569–596 (BAR International Series 232, Oxford, 1985).
2. Knapp, G. *Andean Ecology* (Westview, Boulder, Co, 1993).
3. Kolata, A. *The Tiwanaku* (Blackwell, Cambridge, 1993).
4. Rowe, J. H. in *Handbook of South American Indians* Vol. 2 (ed. Steward, J. H.) 183–330 (Smithsonian Institution, Washington DC, 1947).
5. Steward, J. H. et al. *Irrigation Civilizations* (Pan American Union, Washington DC, 1955).
6. Wittfogel, K. A. *Oriental Despotism* (Yale University, New Haven, CT, 1957).
7. Zimmerer, K. S. *Uso de la tierra y erosión de suelos en la cuenca del Río Calicanto* (CIDRE, Cochabamba, 1991).
8. Zimmerer, K. S. *World Dev.* **21**, 1659–1675 (1993).
9. Zimmerer, K. S. *J. Soil Wat. Conserv.* **49**, 29–34 (1994).
10. GEOBOL-United Nations *Investigaciones de aguas subterráneas en las cuencas de Cochabamba* (GEOBOL-United Nations, Cochabamba, 1978).
11. Amurrio, J. & Espinoza, E. *Clasificación de suelos y tierras con fines de riego a nivel detallado* (CIDRE, Cochabamba, 1991).
12. Bull, W. B. in *Recognition of Ancient Sedimentary Environments* (eds Hamblin, W. K. & Rigby, J. K.) 63–83 (Soc. Econ. Paleo. & Mineral. Spec. Publication 16, Tulsa, OK, 1972).
13. Bull, W. B. *Prog. phys. Geog.* **1**, 222–270 (1977).
14. Harvey, A. M. in *Arid Zone Geomorphology* (ed. Thomas, D. S. G.) 136–158 (Belhaven, London, 1989).
15. Nilsen, T. H. *Modern and Ancient Alluvial Fan Deposits* (Van Nostrand Reinhold, New York, 1985).
16. Baker, V. R. & Ritter, D. F. *Geol. Soc. Am. Bull.* **86**, 975–978 (1975).
17. Moseley, M. E. *The Incas and their Ancestors* (Thames and Hudson, London, 1992).
18. Grieder, T. et al. *La Gálaga, Peru: A Preceramic Culture in Transition* (University of Texas, Austin, 1988).
19. Moseley, M. E. in *Irrigation's Impact on Society* (eds Downing, T. E. & Gibson, M.) 77–82 (University of Arizona, Tucson, 1974).
20. Moseley, M. E. *The Maritime Foundations of Andean Civilization* (Cummings, Menlo Park, 1975).
21. MacNeish, R. S. *The Origins of Agriculture and Settled Life* (University of Oklahoma, Norman, 1992).
22. Pearsall, D. M. in *The Origins of Agriculture* (eds Cowan, C. W. & Watson, P. J.) 173–205 (Smithsonian Institution, Washington DC, 1992).

ACKNOWLEDGEMENTS. I thank V. Holliday, J. Knox, D. Young Holliday, W. M. Denevan, W. Doolittle, G. Knapp, B. Bauer, G. Feinman, S. Lecce, R. Langstroth, S. O. Brooks, and the Centro de Investigación y Desarrollo Regional (CIDRE) for comments, assistance and cooperation. The National Science Foundation and Social Science Research Council funded field research in 1991. The Whitbeck Geography Fund, the Nave Bequest for Latin American and Iberian Studies, and The Land Tenure Center at UW-Madison supported fieldwork in 1990 and 1993 and data analysis in 1994. Research at the Tarata site was performed in affiliation with CIDRE. This study of early irrigation was part of a larger study of the environmental geography of soil erosion and conservation near Tarata and in other Cochabamba highlands.

## Alternative antipredator defences and genetic polymorphism in a pelagic predator–prey system

Luc De Meester<sup>\*†</sup>, Lawrence J. Weider & Ralph Tollrian<sup>\*</sup>

Max-Planck-Institut für Limnologie, Abteilung Ökophysiologie, Postfach 165, 24302 Plön, Germany

<sup>†</sup> Laboratory of Animal Ecology, University of Gent, K. L. Ledeganckstraat 35 9000 Gent, Belgium

**DIEL vertical migration (DVM) of zooplankton is generally considered to be a predator-avoidance strategy: zooplankton migrate to greater depths during the day to reduce their chance of being detected by visual predators (fish)<sup>1</sup>. Both phenotypic plasticity and interpopulational genetic variability in DVM patterns exist in zooplankton<sup>2,3</sup>. We used large indoor mesocosms ('plankton towers'<sup>4</sup>) to study intrapopulational genetic variation for day depth**

**in a *Daphnia hyalina* × *galeata* hybrid population. Clones differing in body size also differed in vertical distribution, with the largest clone residing at the greatest depth during the day. A selection experiment in the presence of fish indicates that alternative anti-predator strategies, which involve a complex association between habitat-selection traits and life-history strategies, might be an important factor underlying intrapopulational genetic polymorphism in zooplankton, through a balancing of fitness effects in the presence of visual predators.**

Predation is often important in the structuring of communities and populations<sup>5,6</sup>. Although theoretical studies<sup>7,8</sup> have attempted to model the impact of predation as a selective force influencing genetic variability in prey populations, there have been few empirical studies of the direct impact of predation on the genetic composition of prey populations<sup>5,9</sup>.

Diel vertical migration (DVM) of zooplankton is generally considered to be a strategy to reduce vulnerability to visual predators<sup>10,11</sup>. All else being equal, larger zooplankton individuals should reside at greater day depths than smaller individuals. This has been observed repeatedly in comparisons of ontogenetic stages of a given species<sup>12,13</sup>, in intraspecific<sup>14,15</sup> and interspecific comparisons<sup>16</sup>.

It has been shown that DVM in zooplankton can be induced by the presence of predator-specific chemicals<sup>2,17–19</sup>, but there is substantial evidence of genetic polymorphism for vertical migration behaviour, at least in the freshwater cladoceran genus *Daphnia*<sup>13,20,21</sup>. In laboratory studies, *Daphnia* clones that inhabit shallow day depths<sup>22</sup> mature at smaller sizes than clones that

<sup>\*</sup> Present addresses: Laboratory of Ecology and Aquaculture, Catholic University of Leuven, Naamsestraat 59, 3000 Leuven, Belgium (L.D.M.); Zoological Institute, Ludwig-Maximilians-Universität Karlstrasse 25 80333 München, Germany (R.T.).

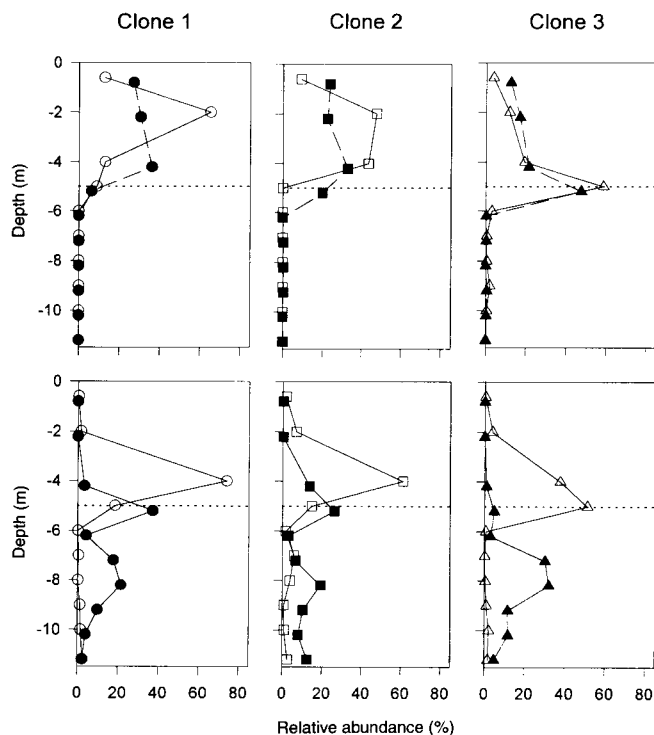


FIG. 1 Vertical distribution of three *Daphnia hyalina* × *galeata* clones in the plankton towers. Top, average vertical distribution of adult females during day (open symbols) and night (filled symbols) in the absence of fish chemicals or fish ( $n=4$  sample series). Bottom: daytime distribution in the presence of fish chemicals (open symbols;  $n=9$ ) and fish (filled symbols;  $n=13$ ). The dotted line indicates depth of thermocline. The plankton towers are large (11.2 m high, cross-section 0.86 m), indoor, twin tower-tank systems<sup>4</sup>. To mimic summer stratification conditions in a temperate lake, a thermocline was established at a depth of 5 m in both towers (epilimnion, 20 °C; hypolimnion, 8 °C). The epilimnion was gently circulated vertically, and food (the unicellular alga *Scenedesmus acutus*) was added daily to a final concentration of 0.8–1.0 mg C l<sup>-1</sup> in the epilimnion (daily feeding was stopped on day 39). At the start of the experiment, both towers were stocked with a mixed population of three naturally coexisting *D. hyalina* × *galeata* clones at a density of 1,000 adult females per clone per tower. All three clones were isolated in July 1994 from the Schöhsee (Plön, Germany) and grown as clonal populations in the laboratory. They were chosen because they could be distinguished by their electromorph patterns<sup>27</sup> at the *PGM* locus, and because they differed in their life histories and vertical distribution (L.D.M. & L.J.W., manuscript in preparation). From day 8 until day 15, one of the towers (day 8–12, tower 2; day 12–15, tower 1) received fish chemicals (10 *Leucaspis delinatus* in a 60-l tub with continuous water flow to the tower and back), while the other tower served as non-fish control<sup>18</sup>. From day 15 onwards, both towers received fish chemicals. On day 24, 10 *L. delinatus* (size 3–4 cm) were released into tower 1. On day 27, 50 *L. delinatus* (size 2–4 cm) were released into tower 2. During the sampling bouts (during the day at 12:30 and night at 22:30), *Daphnia* were collected simultaneously at depths of 0.6, 2.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0 and 11.0 m, using pumps and flow-through traps<sup>4</sup>. If available, 40 adult females were analysed from each epilimnetic and metalimnetic depth, and 20 from each hypolimnetic depth. Individuals ( $n=8,500$ ) were measured and their clutch sizes recorded, and they were frozen for electrophoresis<sup>27</sup>. Overall, there are significant differences in average depth between clones 1 and 3 during the day, between all three genotypes during the night, and between day and night samples of each of the three clones (Wilcoxon,  $n=24-26$ ,  $P<0.01$ , Bonferroni correction<sup>28</sup>).

reside deeper during the day<sup>23</sup>. There is a positive correlation between adult body size and number, size and starvation resistance of offspring<sup>24,25</sup>. Larger individuals are, however, more vulnerable to visual predation than smaller ones. Migrating to greater depths during the day can reduce vulnerability to visual predation, but it entails a metabolic cost<sup>11</sup>. As a result of these

conflicting pressures, multiple optima may exist, with non-migrating small individuals (or clones) and larger individuals (or clones) that migrate to greater depths during the day having essentially equal fitnesses.

We tested the hypothesis that differences in day depth might compensate for differences in body size with respect to the persistence of different clones under visual predation. To do this we monitored changes in clonal frequencies in mixed populations before and after the introduction of fish. Mixed populations of three naturally coexisting *Daphnia hyalina* × *galeata* hybrid clones, isolated from the Schöhsee (Germany) and characterized by a different size at maturity, were established in large, indoor, plankton towers (Fig. 1). The clones exhibited different vertical distributions in the towers, with the largest clone residing deeper in the water column than the two smaller clones (Fig. 1). During the day and in the absence of fish chemicals, most adults of clones 1 and 2 were found at 2–4 m, whereas most adults of the largest clone, clone 3, were observed at 5 m (near the thermocline). Clones 1 and 2 were distributed throughout the epilimnion during the night, whereas the vertical distribution of clone 3 did not differ appreciably between day and night. In the presence of fish chemicals, and after stocking the towers with fish, adults of clone 3 tended to reside deeper than those of the other two clones (Fig. 1).

Adults of clone 3 were larger on average than those of clones 1 and 2 (Fig. 2a). However, even though clones 1 and 3 differed

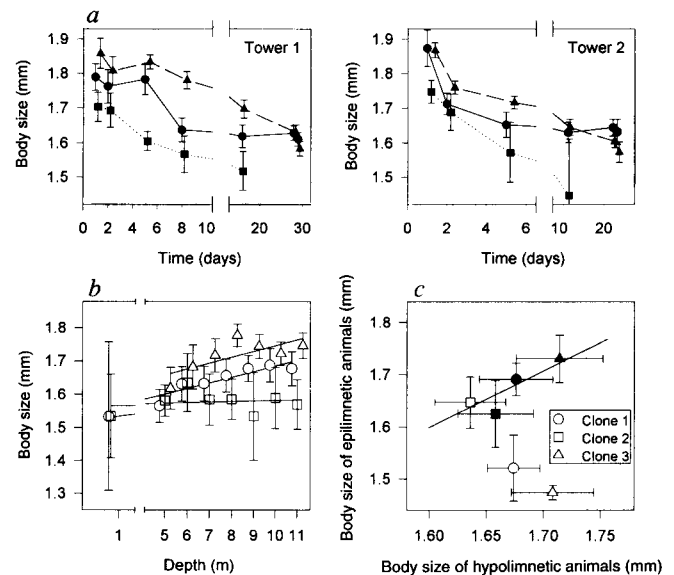


FIG. 2 Body size of clones 1 (circles), 2 (squares) and 3 (triangles) in the plankton towers. a, Changes in mean ( $\pm 2$  s.e.) size of adult females of the three clones in the plankton towers after the addition of fish (day 1 was the first day with fish present). b, Mean ( $\pm 2$  s.e.) body size of adult females of the three clones plotted against the depth at which they were caught. Pooled daytime samples from both towers, taken on three successive sampling days in the presence of fish; few individuals of clone 1 ( $n=3$ ) and clone 2 ( $n=9$ ) were caught in the epilimnion, and none of clone 3 were detected. These data were grouped into three depth strata (epi- plus metalimnion, 0–5 m; upper hypolimnion, 6–7 m; and lower hypolimnion, 8–11 m). Slopes in clones 1 and 3 are homogeneous ( $P=0.999$ ), whereas the slopes are heterogeneous ( $P=0.013$ ) when clone 2 data are included (ANCOVA). c, Mean size ( $\pm 1$  s.e.) of adults in epilimnion plotted against size of adults in the hypolimnion, for day (open symbols) and night (filled symbols) samples. Clonal means were calculated using mean epilimnetic or hypolimnetic values for a given sample series (with values only included if based on 10 or more animals;  $n=4$  series for daytime epilimnetic individuals for all three clones and night-time hypolimnetic clone 2;  $n=7-11$  series for all other data; large error bars are due to changes in body size with time; see Fig. 2a). The line connects points of equal epi- and hypolimnetic body size.

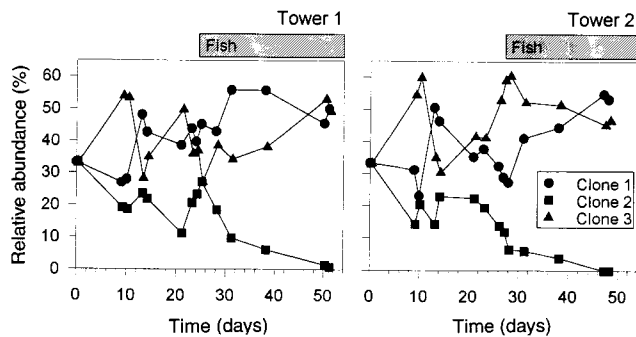


FIG. 3 Relative abundance (based on night-time sample series) of clones 1 (circles), 2 (squares) and 3 (triangles) in the two plankton towers. Frequencies (0.333) were equal at the start of the experiment. The shaded bars above the graphs show when fish were present. Birth rates<sup>29</sup> were calculated for a short period after fish were added (day 28–30), and were compared with estimates of  $r$  obtained from population growth rates during the same period. Approximately 1,700 and 500 *Daphnia* per fish per day were eaten in towers 1 and 2, respectively.

substantially in size, they both showed a very similar relationship between size and depth: adults of these clones that ventured into the epilimnion during the day were invariably small (Fig. 2b). During the day, adults of clones 1 and 3 that remained in the epilimnion were much smaller than animals of the same clones that resided in the hypolimnion (Fig. 2c; there were no size differences between epi- and hypolimnetic animals during the night). For clone 2 there was no such difference in size between epilimnetic and hypolimnetic individuals (Fig. 2c). As a result, the average size of the epilimnetic adult females during the day was larger for clone 2 than for the other clones, even though clone 2 was the smallest clone on average (Fig. 2a).

Overall, changes in clonal frequencies were similar in both towers (Fig. 3). Before fish were added the relative abundances of clones 1 and 3 converged at around 40%, with the relative abundance of clone 2 being about 20%. After stocking the towers with fish, the relative abundance of clone 2 (the only clone where large animals remained in the epilimnion during the day) was gradually reduced until near-extinction (0% in tower 2, and <1% in tower 1), whereas clones 1 and 3 continued to coexist, being at frequencies of approximately 50% at end of the experiment (Fig. 3).

We conclude that clones 1 and 3, which differ strongly in both DVM and life-history strategies, have similar fitnesses in the presence of planktivorous fish. Clone 2 may represent a third strategy, with individuals relying on their small size to remain longer in the metabolically more profitable epilimnion. In our experiment, clone 2 clearly had a lower fitness than the other two clones. However, in natural environments the relative fitness of clones with different migration strategies might depend critically on predation pressure and lake stratification. Under lower predation pressure than that established in the towers, it might be advantageous to remain in the epilimnion, irrespective of body-size characteristics.

This study shows that an association between behavioural (habitat selection) and life-history traits may result in alternative strategies, which in turn result in equal fitnesses for different genotypes when exposed to predation pressure. Such a balancing of fitness effects can perhaps explain the maintenance of genetic polymorphism for DVM in *Daphnia*, and possibly other zooplankton as well. More generally, our results illustrate the context under which complex strategies can evolve, and might provide a mechanism to promote the coexistence of genotypes in natural populations<sup>26</sup>. □

- Lampert, W. *Funct. Ecol.* **3**, 21–27 (1989).
- Huntley, M. & Brooks, E. R. *Mar. Biol.* **71**, 23–31 (1982).
- Leibold, M. A., Tessier, A. J. & West, C. T. *Evolution* **48**, 1324–1332 (1994).
- Guisande, C., Duncan, A. & Lampert, W. *Oecologia* **87**, 357–359 (1991).
- Leibold, M. A. & Tessier, A. J. *Oecologia* **86**, 342–348 (1991).
- Lampert, W. & Sommer, U. *Limnol. Oekologie* (Thieme, New York, 1993).
- Ringelberg, J. J. *Plankton Res.* **13**, 83–89 (1991).
- Loose, C. J. *Arch. Hydrobiol. Beih. Ergebn. Limnol.* **39**, 29–36 (1993).
- De Meester, L. *Ecology* **74**, 1467–1474 (1993).
- Weider, L. J. *Limnol. Oceanogr.* **29**, 225–235 (1984).
- De Meester, L. *Hydrobiologia* **225**, 217–227 (1991).
- De Meester, L. *Arch. Hydrobiol. Beih. Ergebn. Limnol.* **39**, 137–155 (1993).
- De Meester, L. *Oecologia* **97**, 333–341 (1994).
- Ebert, D. *Arch. Hydrobiol. (suppl.)* **90**, 453–473 (1993).
- Tessier, A. J. & Consolatti, N. L. *Oikos* **58**, 269–276 (1989).
- Hedrick, P. W. A. *Rev. Ecol. Syst.* **17**, 535–566 (1986).
- Hebert, P. D. N. & Beaton, M. J. *Methodologies for Allozyme Analysis using Cellulose Acetate Electrophoresis* (Helena Laboratories, Beaumont, TX, 1989).
- Rice, W. R. *Evolution* **43**, 223–225 (1989).
- Paloheimo, J. E. *Limnol. Oceanogr.* **19**, 692–694 (1974).

ACKNOWLEDGEMENTS. We thank W. Lampert for encouragement and support, P. Spaak and three anonymous reviewers for providing comments on the manuscript, and E. Geissler, I. Senkpiel and S. Tielke for help with the electrophoresis and maintenance of the towers. L.D.M. is a postdoctoral researcher with the National Science Fund (Belgium). A postdoctoral fellowship from the Max Planck Society to L.D.M., and financial support by the FKFO (Belgium) are gratefully acknowledged.

## Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids

Randall V. Collura & Caro-Beth Stewart\*

Department of Biological Sciences, University at Albany, State University of New York, Albany, New York 12222, USA

USING oligonucleotide primers designed to match conserved regions of mammalian mitochondrial DNA (mtDNA)<sup>1</sup>, we have amplified and sequenced two divergent cytochrome *b* nuclear pseudogenes from orangutan cellular DNA. Evolutionary analysis suggests that a nuclear transfer occurred about 30 million years ago on the lineage leading to the catarrhines (Old World monkeys and hominoids), and involved a long (at least 3 kilobases), probably damaged, piece of mtDNA. After this transfer, the pseudogene duplicated, giving rise to the two copies that are probably present in all hominoids, including humans. More recent transfers involving the entire cytochrome *b* gene have also occurred in the Old World monkeys. Such nuclear copies of mtDNA can confound phylogenetic and population genetic studies<sup>2–4</sup>, and be an insidious source of DNA contamination of 'ancient'<sup>3,5</sup> and forensic DNA. Indeed, contamination with these anciently transferred human pseudogenes<sup>5</sup> is almost certainly the source of the cytochrome *b* sequences recently reported from 'dinosaur bone DNA'<sup>6</sup>.

\* To whom correspondence should be addressed.

Received 5 June; accepted 2 October 1995.

- Stich, H.-B. & Lampert, W. *Nature* **293**, 396–398 (1981).
- Neill, W. E. *Nature* **345**, 524–526 (1990).
- Neill, W. E. *Nature* **356**, 54–57 (1992).
- Lampert, W. & Loose, C. J. *Arch. Hydrobiol.* **126**, 53–66 (1992).
- Endler, J. A. *Natural Selection in the Wild* (Princeton Univ. Press, 1986).
- Kerfoot, W. C. & Sih, A. (eds) *Predation, Direct and Indirect Impacts on Aquatic Communities* (Univ. Press New England, Hanover, 1987).
- So, J. W. H. & Freedman, H. I. *Bull. math. Biol.* **48**, 469–484 (1986).
- So, J. W. H. *J. Aust. math. Soc. B* **31**, 347–365 (1990).
- Pijanowska, J., Weider, L. J. & Lampert, W. *Oecologia* **96**, 40–42 (1993).
- Zaret, T. M. & Suffern, J. S. *Limnol. Oceanogr.* **21**, 804–813 (1976).

Amplification by the polymerase chain reaction (PCR)<sup>7</sup> of mtDNA followed by sequencing of the PCR product<sup>8</sup> is widely used today in molecular evolutionary studies<sup>1</sup> and has been advocated for forensics<sup>9</sup>. Direct sequencing of the PCR product usually yields accurate mtDNA sequence, but the reliability of this method depends on the specificity of the primers for mtDNA<sup>2,4</sup>.

We have been using this approach to determine the mtDNA phylogeny of the colobine Old World monkeys. Using primers designed to match conserved transfer RNA genes flanking the cytochrome *b* gene, we have amplified and sequenced this entire 1.2-kilobase (kb) region from cellular DNAs of numerous primate species, including several hominoids for controls (Fig. 1). For most species, sequences obtained directly from the PCR products were unambiguous and predicted full-length functional proteins.

In contrast, direct sequencing of the single-band PCR product amplified from Sumatran orangutan cellular DNA suggested the presence of multiple sequences. Amplifications using mtDNA purified in a caesium chloride gradient<sup>10</sup> still showed multiple sequences. (Retrospective examination of primer annealing sites revealed that one of the primers has more mismatches, including the 3' nucleotide, to the orangutan mitochondrial sequence than to the nuclear copies. This illustrates that the specificity of the PCR can override attempts to purify mtDNA if the primers

anneal preferentially to even minor DNA contaminants<sup>2</sup>.) We made several attempts to amplify the functional mitochondrial copy cleanly by using different PCR conditions and primer sets designed to longer or shorter segments of mtDNA. Only one such product yielded an unambiguous sequence, but this 3-kb fragment clearly did not code for a functional cytochrome *b* protein, as it contained frameshift mutations and stop codons (not shown). By comparing conserved regions of the primate mitochondrial cytochrome *b* sequences to this pseudogene ('orangutan nuclear I'), internal primer pairs were designed that cleanly amplified the Sumatran orangutan mitochondrial gene. In addition, a second cytochrome *b* pseudogene ('orangutan nuclear II') was identified from the same individual (Fig. 1). These two pseudogenes share many mutations, including a frameshift deletion and two mtDNA stop codons (not shown). The entire mitochondrial genomes of several hominoids, including Bornean orangutan, have recently been sequenced<sup>11</sup> and contain no such pseudogenes. Thus, these sequences are almost certainly of nuclear origin (see also ref. 5) and, as they contain many nuclear stop codons, are unlikely to be transcribed.

Indeed, the pattern of molecular evolution between these two orangutan pseudogenes is consistent with neutral evolution in the catarrhine nucleus<sup>12,13</sup>, but not with mtDNA evolution<sup>11,14</sup>. These pseudogenes are too divergent (8% sequence difference) to be neutral alleles at the same locus. Taken together, this sug-

|                      | 15624 | 15634 | 15644 | 15654 | 15664 | 15674 | 15684 | 15694 | 15704 | 15714 | 15724 | 15734 | 15744 | 15756 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |    |
|----------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
| 1 Human              | CC    | CTA   | TTA   | CTA   | TCC   | ATC   | ATC   | CTA   | GCA   | AGA   | ATC   | CCC   | ATC   | CTC   | CAT | ATA | TCC | -   | AAA | CAG | CAA | AGC | ATA | ATA | TTC | CGC | CCA | CTA | AGC | CAA | TCA | TTT | TAT | TGA | CTC | CTA | GCC | ACA | GAC | CTC | CTC | ATT | CTA | ACC | TG  |     |     |    |
| 2 Chimpanzee         | CC    | CTA   | TGA   | CTA   | TCC   | ATC   | ATC   | CTA   | GCA   | AGA   | ATC   | CCC   | ATC   | CTC   | CAT | ATA | TCC | -   | AAA | CAG | CAA | AGC | ATA | ATA | TTC | CGC | CCA | CTA | AGC | CAA | TCA | TTT | TAT | TGA | CTC | CTA | GCC | ACA | GAC | CTC | CTC | ATT | CTA | ACC | TG  |     |     |    |
| 3 Gorilla            | CC    | CTA   | TGA   | CTA   | TCC   | ATC   | ATC   | CTA   | GCA   | AGA   | ATC   | CCC   | ATC   | CTC   | CAT | ATA | TCC | -   | AAA | CAG | CAA | AGC | ATA | ATA | TTC | CGC | CCA | CTA | AGC | CAA | TCA | TTT | TAT | TGA | CTC | CTA | GCC | ACA | GAC | CTC | CTC | ATT | CTA | ACC | TG  |     |     |    |
| 4 Orangutan*         | CC    | CTC   | ATG   | CTA   | TCC   | ATC   | ATC   | CTA   | GCA   | AGA   | ATC   | CCC   | ATC   | CTC   | CAT | ATA | TCC | -   | AAA | CAG | CAA | AGC | ATA | ATA | TTC | CGC | CCA | CTA | AGC | CAA | TCA | TTT | TAT | TGA | CTC | CTA | GCC | ACA | GAC | CTC | CTC | ATT | CTA | ACC | TG  |     |     |    |
| 5 Gibbon*            | CC    | CTC   | ATG   | CTA   | TCC   | ATC   | ATC   | CTA   | GCA   | AGA   | ATC   | CCC   | ATC   | CTC   | CAT | ATA | TCC | -   | AAA | CAG | CAA | AGC | ATA | ATA | TTC | CGC | CCA | CTA | AGC | CAA | TCA | TTT | TAT | TGA | CTC | CTA | GCC | ACA | GAC | CTC | CTC | ATT | CTA | ACC | TG  |     |     |    |
| 6 Colobus Monkey*    | CC    | CTC   | ATG   | CTA   | TCC   | ATC   | ATC   | CTA   | GCA   | AGA   | ATC   | CCC   | ATC   | CTC   | CAT | ATA | TCC | -   | AAA | CAG | CAA | AGC | ATA | ATA | TTC | CGC | CCA | CTA | AGC | CAA | TCA | TTT | TAT | TGA | CTC | CTA | GCC | ACA | GAC | CTC | CTC | ATT | CTA | ACC | TG  |     |     |    |
| 7 Rhesus Monkey*     | CA    | CTC   | TTC   | CTA   | TCA   | ATC   | CTC   | ATC   | GCA   | GCA   | ATC   | CCC   | ATC   | CTC   | CAC | AAA | TCC | -   | AAA | CAG | CAA | AGT | ATA | AGA | TTC | CGC | CCA | CTC | AGC | CAA | TTC | CTG | TTC | TTC | CTC | CTC | CTA | ACT | ACA | ACT | ATC | ATC | ACC | CTC | ACC | TG  |     |    |
| 8 Human Nuclear*     | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | GCA   | GTG   | GTG   | ATC   | GCC   | CTC | CAC | AAA | TGG | -   | C   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | CGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |    |
| 9 Orang Nuclear I*   | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | GCA   | GTG   | ATC   | ATC   | GCA   | CTC | CAC | AAA | TCC | -   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | TGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |     |    |
| 10 Orang Nuclear II* | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | GCA   | GTG   | ATC   | ATC   | GCA   | CTC | CAC | AAA | TCC | -   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | TGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |     |    |
| 11 Gibbon Nuclear*   | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | GCA   | GTG   | ATC   | ATC   | GCA   | CTC | CAC | AAA | TCC | -   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | CGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |     |    |
| 12 Gorilla Nuclear*  | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | GCA   | GTG   | ATC   | ATC   | GCA   | CTC | CAC | AAA | TCC | -   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | CGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |     |    |
| 13 Bonobo Nuclear*   | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | GCA   | GTG   | ATC   | ATC   | ATC   | GCA   | CTC | CAC | AAA | TCC | -   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | TGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |     |    |
| 14 Chimp Nuclear*    | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | GCA   | GTG   | ATC   | ATC   | ATC   | GCA   | CTC | CAC | AAA | TCC | -   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | TGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |     |    |
| 15 Bone 4-37         | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | TTC   | GTG   | GTG   | ATC   | GCC   | CTC | CAC | AAA | TCC | -   | C   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | CGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |    |
| 16 Bone 31-44        | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | TTC   | GTG   | GTG   | ATC   | GCC   | CTC | CAC | AAA | TCC | -   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | CGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |     |    |
| 17 Bone 2-61         | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | TTC   | GGA   | ATC   | ATC   | ATC   | CTA | CTC | CAC | AAA | TCC | -   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | CGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |    |
| 18 Bone 2-18         | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | TTC   | GGA   | ATC   | ATC   | ATC   | CTA | CTC | CAC | AAA | TCC | -   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | CGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |    |
| 19 Bone 20-61        | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | TTC   | GTG   | ATC   | ATC   | ATC   | CTA | CTC | CAC | AAA | TCC | -   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | CGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |    |
| 20 Bone 2-37         | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | TTC   | GTG   | ATC   | ATC   | ATC   | CTA | CTC | CAC | AAA | TCC | -   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | CGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |    |
| 21 Bone 1-37         | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | TTC   | GTG   | ATC   | ATC   | ATC   | CTA | CTC | CAC | AAA | TCC | -   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | CGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |    |
| 22 Bone 5-37         | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | TTC   | GTG   | ATC   | ATC   | ATC   | CTA | CTC | CAC | AAA | TCC | -   | C   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | CGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG |
| 23 Bone 6-37         | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | TTC   | GTG   | ATC   | ATC   | ATC   | CTA | CTC | CAC | AAA | TCC | -   | AAA | CAG | CAA | AGC | ATC | ATA | TTC | CGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | GAC | CTA | TTC | ACA | CTC | ACA | TG  |    |
| 24 New World Monkey* | CC    | CTT   | GGA   | TGA   | TCC   | ATC   | CTC   | ATC   | CTA   | GTC   | GTG   | GTG   | ATC   | GCC   | CTC | CAC | CTG | TCA | -   | AAC | CAG | CAA | AGT | ATG | ATA | TTC | CGC | CCA | CTA | AGT | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | AAC | CTA | CTT | ACA | CTC | ACA | TG  |     |    |
| 25 Lemur*            | CC    | CTA   | GTA   | TTC   | TTC   | ATC   | ATC   | ATC   | CTA   | ATT   | CT    | GTA   | ATC   | ATC   | ACC | AGT | CTA | CAC | GGC | -   | AAA | CAG | TAT | ATG | GTA | TTC | CGC | GGC | CTC | ACC | CAA | TAC | TGA | TTC | TGA | ATC | CTA | AGA | GAT | TGA | ATT | ATT | CTT | ACA | TG  |     |     |    |
| 26 Fin Whale         | CC    | CTA   | GTA   | CTC   | TCA   | ATC   | ATC   | CT    | GCC   | TTC   | ATC   | ATC   | ATC   | ATC   | CTA | CTC | CAC | AGA | TCC | -   | AAT | CAG | ACA | AGC | ATA | ATA | TTC | CGC | GGC | CTT | AGC | GAG | TTC | TTC | TTC | GTA | GTC | CTA | GCA | GAT | CTA | CTC | ATC | CTA | ACC | CTA | TG  |    |
| 27 Olive Vireo       | CC    | CTA   | GCC   | GCC   | TTC   | ATC   | CTC   | CTA   | ATC   | ATC   | ATC   | ATC   | ATC   | ATC   | CTA | CTC | CAC | AGA | TCC | -   | AAA | CAG | AGC | ATA | ATA | TTC | CGC | GGC | CTC | ACC | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | AAC | CTA | CTT | ACA | CTC | ACA | TG  |     |    |
| 28 Blue Jay          | CC    | CTA   | GCC   | GCC   | TTC   | ATC   | CTC   | CTA   | ATC   | ATC   | ATC   | ATC   | ATC   | ATC   | CTA | CTC | CAC | AGA | TCC | -   | AAA | CAG | AGC | ATA | ATA | TTC | CGC | GGC | CTC | ACC | CAA | TAT | CTG | TCC | TGA | ATC | CTA | GTC | GGC | AAC | CTA | CTT | ACA | CTC | ACA | TG  |     |    |
| 29 Least Flycatcher  | CG    | CTC   | GCC   | GCC   | TTC   | GTC   | CTC   | GTC   | CTC   | GTC   | CTC   | GCC   | CTC   | GCC   | CTC | CAC | ATA | TCA | -   | AAA | CAG | AGC | ATA | ACC | TTC | GTC | GGC | CTC | ACC | CAA | CTA | TTC | TTC | TTC | TGA | ATC | CTA | GTC | GGC | AAC | CTA | CTC | GTT | CTT | ACA | TG  |     |    |
| 30 Chachalaca        | CA    | CTA   | GCA   | GCT   | CTA   | GTA   | ATC   | CTC   | TTC   | CTT   | ATC   | CTC   | TTC   | CTG   | CTG | CAC | AAA | TCC | -   | AAA | CAG | ACA | ATA | AGA | TTC | CGC | GGC | CTC | ACC | CAA | CTC | CTA | TTC | TTC | TGA | ATC | CTA | GTC | GGC | AAC | CTA | CTC | ATC | ATC | ACA | TG  |     |    |
| 31 Muscovy Duck      | CA    | CTA   | GCC   | GCC   | CTC   | GTC   | ATC   | CTA   | TTC   | TTC   | CTC   | CTC   | CTC   | CTC   | CTC | CAC | AGA | TCC | -   | AAA | CAG | ACA | ATG | AGA | TTC | CGC | GGC | CTC | ACC | CAA | CTC | CTA | TTC | TTC | TGA | ATC | CTA | GTC | GGC | AAC | CTA | CTC | ATC | ATC | ACA | TG  |     |    |
| 32 Xenopus           | CC    | CTA   | GTC   | GTC   | CTC   | ATC   | CTC   | CTA   | TTC   | CTC   | CTC   | CTC   | CTC   | CTC   | CTC | CAC | AGA | TCC | -   | AAA | CAG | ACA | ATA | AGA | TTC | CGC | GGC | CTC | ACC | CAA | CTC | CTA | TTC | TTC | TGA | ATC | CTA | GTC | GGC | AAC | CTA | CTC | ATC | ATC | ACA | TG  |     |    |
| 33 Carp              | CA    | CTA   | TTC   | TCC   | ATT   | CTG   | GTA   | ATA   | ATA   | GTA   | GTA   | CTA   | CTA   | CTA   | CAC | ACC | TCC | -   | AAA | CAG | ACA | CTA | AGA | TTC | CGC | GGC | CTC | ACC | CAA | CTC | CTA | TTC | TTC | TGA | ATC | CTA | GTC | GGC | AAC | CTA | ATT | ATC | ATC | ACA | TG  |     |     |    |

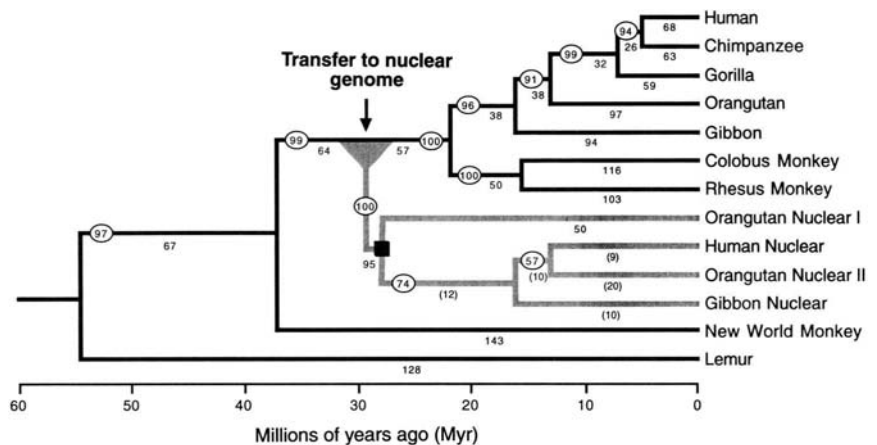
FIG. 1 Partial sequences of primate and other vertebrate mitochondrial cytochrome *b* genes compared to sequences of primate nuclear pseudogenes and sequences from 'dinosaur bone DNA'. The sequences are divided into codons and numbered according to the complete human mitochondrial genome<sup>24</sup>. The human (GenBank J01415), chimpanzee (DDBJ D38113), gorilla (DDBJ D38114), fin whale (EMBL X61145), chicken (EMBL X52392), blue jay (EMBL X74258), muscovy duck (GenBank L08385), chachalaca (GenBank L08384), least flycatcher (EMBL X74251), olive vireo (EMBL X74260), *Xenopus* frog (EMBL X02890), and carp (EMBL X61010) cytochrome *b* sequences shown have all been previously published. The 'dinosaur bone' sequences are as presented in ref. 6. All other sequences (indicated by asterisks) were produced in the present study. Only those regions that align to the 'dinosaur' sequence are shown; the complete alignment of the full-length sequences is available from the authors. These sequences will be deposited into GenBank under accession numbers U38262–U38274. Sequences from additional species will be presented elsewhere.

**METHODS.** Whole cellular DNA was extracted from EDTA anticoagulated blood<sup>25</sup> or frozen tissue samples<sup>10</sup>. The full-length cytochrome *b* gene (1,140 bp) was amplified and sequenced from agile gibbon (*Hylobates agilis*; 'Susie' at Houston Zoo), Old World black-and-white colobus monkey (*Colobus guereza*; '073228' at St Louis Zoo), Old World rhesus monkey (*Macaca mulatta*; purified mtDNA, from T. Disotelli), New World squirrel monkey (*Saimiri sciureus*; purified cellular DNA, from the late A. C. Wilson), and prosimian ring-tailed lemur (*Lemur catta*; 'Brennus'

at Duke University Primate Center) using primers that anneal to flanking tRNAs (L14724, 5'-CGAGATCTGAAAAACCATCGTTG-3'; H15916, 5'-AACTGCAGTCACTCCGGTTTACAAGA-3'). Sumatran orangutan (*Pongo pygmaeus*; 'Ini' at Yerkes) amplification primers were as follows: mitochondrial cytochrome *b* (L14724, 5'-CGAGATCTGAAAAACCATCGTTG-3'; H15924, 5'-GAGGGTCTTCATTTTCAGG-3'); 'orangutan nuclear I' (L13048, 5'-GGTCTCCACCCCTGACTCCCTCAGCCATAG-3'; H15989, 5'-GAATAGTTTAAATTAGAATCTCAGCTTTGGG-3'); 'orangutan nuclear II' (L14888, 5'-GCTTAATCTCCAAATCACCACAG-3'; H15923, 5'-GGGGA-CTCTCCATTTCTGCT-3'). Partial cytochrome *b* pseudogene sequences (about 420 bp) were amplified from human (*Homo sapiens*; 'RVC'), chimpanzee (*Pan troglodytes*; 'Beleka' at Yerkes), bonobo (*Pan paniscus*; 'Kanzi' at Yerkes), gorilla (*Gorilla gorilla*; 'Banga' at Yerkes) and agile gibbon (*Hylobates agilis*; 'Susie' at Houston Zoo) using primers designed to match orangutan nuclear II (L15415, 5'-GACAAAATC-ACCTTCCACCCCTAC-3'; H15835, 5'-CGGTGAGT



FIG. 2 Phylogenetic tree of selected primate cytochrome *b* sequences (in black) showing placement of nuclear pseudogene sequences (in grey). Evolutionary trees were inferred using the full-length mitochondrial cytochrome *b* gene sequences aligned with the 1,108-bp orangutan nuclear I, the 997-bp orangutan nuclear II, and the 420-bp human and gibbon pseudogene sequences (Fig. 1). Unsequenced ends of the pseudogenes were coded as 'missing data'. Trees were inferred using a variety of parsimony, distance, and maximum likelihood approaches in PHYLIP<sup>26</sup>, PAUP<sup>27</sup> and a test version of PAUP\* (D. L. Swofford, personal communication). The tree shown here is the shortest nucleotide parsimony tree found. It requires 3,278 steps, with transversions weighted 4 times transitions; each insertion/deletion was counted as a single event (weighted 4 times a transition), regardless of the number of nucleotides involved. This tree is rooted with fin whale (not shown) as an outgroup. 2,000 bootstrap replications were performed using a heuristic algorithm; these values are shown in ovals. The average number of unweighted nucleotide substitutions per lineage (ignoring gaps) was calculated using MacClade<sup>28</sup>; these rounded values are shown below the lineages. Values in parentheses are based on partial sequences. All tree inference methods gave values of roughly similar proportion. Divergences are assumed to represent speciation events, except for the transfer (indicated by the arrow) and the pseudogene duplication (indicated by the square). Dates of divergence of the major primate lineages are average estimates from the fossil record<sup>29</sup>; those of the great ape lineages were calculated from their complete mtDNA sequences<sup>11</sup>. Using these latter dates and the full-length cytochrome *b* sequences, we calculate that colobus (a colobine Old World monkey) and rhesus (a cercopithecine Old World monkey) diverged about 15 Myr ago, and that gibbon diverged from the great apes about 16 Myr ago. Based on the average number of substitutions before and after the pseudogene lineage attaches to the ancestral catarrhine lineage, we estimate the transfer event at about 30 Myr (approximate range, 27–33 Myr, given other possible branch lengths found by different methods). Based on the rate of nucleotide substitution that we calculated for the human and gibbon pseudogenes of



$1.5 \times 10^{-9}$  substitutions per site per year (a value consistent with the most recent estimates of the neutral substitution rate between hominoid orthologues<sup>12</sup>), we estimate that the duplication of the two orangutan pseudogenes took place ~26–30 Myr ago (range, 22–45 Myr). Evolutionary trees were also inferred by the Fitch–Margoliash, neighbour-joining, and maximum likelihood methods using a variety of nucleotide substitution models. When transversions were weighted more heavily than transitions, a single tree was found by all distance methods. The only way in which this tree differs from the shortest parsimony tree is that it shows the human nuclear pseudogene as more closely related to the gibbon nuclear pseudogene than to the orangutan nuclear II pseudogene. The parsimony tree of this topology requires 3,280 steps, and is the second-shortest tree found by PAUP<sup>27</sup>. If this alternative branching order is correct, a second gene duplication event must be postulated. The maximum likelihood tree is also nearly identical to the shortest parsimony tree, except that the two orangutan pseudogenes are most closely related; this tree would require a minimum of three pseudogene duplications. The parsimony tree of this topology requires 3,283 steps. Given the closeness of these alternative trees and the low bootstrap numbers within the pseudogene clade, the duplication order should be considered tentative.

gests that they are most likely to be gene duplicates. The primate cytochrome *b* sequences we generated allow us to trace the evolutionary history of these pseudogenes by phylogenetic analysis.

Evolutionary trees inferred from the mitochondrial cytochrome *b* genes (Fig. 2, solid lines) are congruent with the known branching order of the major primate lineages, including the hominoids<sup>11</sup>, affirming that cytochrome *b* is a reliable phylogenetic marker in these primates. Trees inferred by all methods (Fig. 2) show the orangutan pseudogenes forming a separate clade which attaches to the mtDNA phylogeny about halfway along the ancestral catarrhine lineage (that is, after the divergence from the New World monkey lineage, but before the divergence of the Old World monkey and hominoid lineages). This supports the hypothesis that these pseudogenes are duplicates of a single transferred segment of mtDNA, and suggests that this transfer occurred approximately 30 million years ago in an ancestor of the modern catarrhines. Similar transfers and duplications of large segments of mtDNA have been documented along the cat<sup>15</sup> and rat<sup>16</sup> lineages.

Based on the catarrhine neutral substitution rate<sup>12</sup>, we estimate that this duplication took place within a few million years after the transfer, probably before the divergence of the Old World monkey and hominoid lineages (Fig. 2). Indeed, using primers designed to amplify a 420-base-pair (bp) region of orangutan nuclear II, but to exclude amplification of the other known sequences, we characterized cytochrome *b* pseudogenes in other hominoid (examples shown in Fig. 1) and Old World monkey species (not shown). Trees inferred by all methods group the human and gibbon pseudogenes together with the orangutan nuclear sequences (Fig. 2, grey lines). Whereas bootstrap sup-

port for the placement of this pseudogene clade is very high (99–100%), the number of substitutions assigned to the short lineage after the transfer but before the duplication (Fig. 2) is several times higher than can be accounted for solely by neutral evolution in a catarrhine nucleus. Some of these substitutions are consistent with mtDNA evolution, so they could be due to ancient polymorphism or to incorrect reconstructions; however, many are not consistent with a functional cytochrome *b* gene. A reasonable explanation for this latter category of mutation is that the mtDNA fragment that integrated into the nucleus was previously damaged<sup>17</sup> and not correctly repaired.

Gorilla, bonobo, chimpanzee and colobus monkey have cytochrome *b* pseudogenes that apparently were transferred to the nucleus along the ancestral catarrhine lineage; these sequences are not shown in the trees because it is not certain whether they are products of additional duplications or separate transfers. Furthermore, Old World monkeys have numerous full-length cytochrome *b* pseudogenes that were transferred more recently, as they branch with Old World monkey mitochondrial cytochrome *b* sequences in evolutionary trees (not shown). Similar groups of 12S ribosomal DNA pseudogenes have been recently reported in the catarrhines<sup>3</sup>. It is thus clear that catarrhines have many mtDNA-like pseudogenes which were transferred to the nucleus at different times in evolutionary history<sup>3,17,19</sup>. Indeed, hybridization studies using human nuclear DNA libraries suggested that hundreds of copies of mtDNA-like segments, some quite long, are present in the human nuclear genome<sup>19</sup>. Unravelling the complex evolutionary history of these primate pseudogenes will require extensive molecular studies to characterize the sites of insertion, coupled with careful phylogenetic analyses.

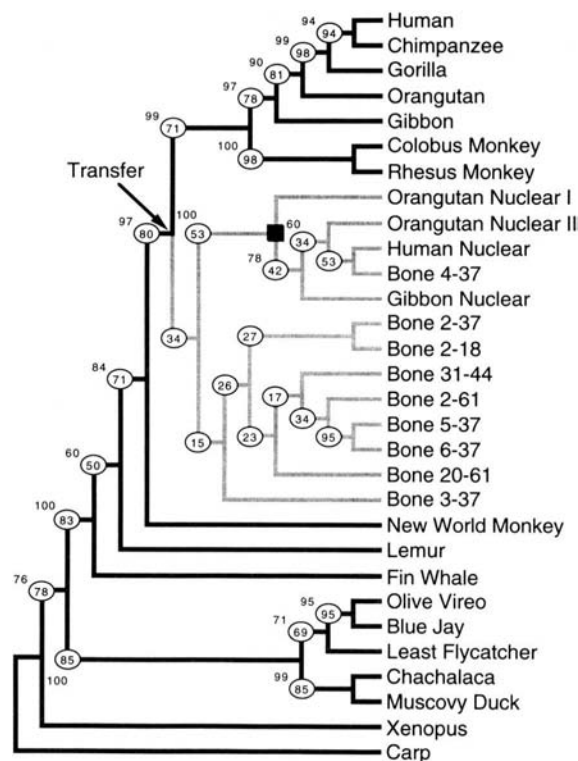


FIG. 3 Phylogenetic tree of selected vertebrate cytochrome *b* sequences (in black) showing relationship of the reported 'dinosaur bone' sequences to the nuclear pseudogenes (in grey). Evolutionary trees were inferred as for Fig. 2. Included here are the 83–134-bp 'dinosaur bone' sequences<sup>6</sup> (Fig. 1) and full-length sequences from selected additional species. The tree shown is the shortest nucleotide parsimony tree (6,587 steps) found by repeated heuristic searches using the computer program PAUP<sup>27</sup>. This tree is rooted with a fish (carp) sequence as outgroup to the tetrapods. Dates of divergence are not implied on this tree. 500 bootstrap replications were performed using all taxa shown; values are shown in ovals. 1,000 bootstrap replications were performed using all taxa except the 'dinosaur bone' sequences; these values are shown beside the ovals. In most cases, the addition of the short 'dinosaur bone' sequences reduced the bootstrap support for individual clades. Trees that added only the 'bone 4-37' sequence showed high bootstrap support (95%) for its branching with the human nuclear pseudogene sequence (not shown). Evolutionary trees were also inferred by the Fitch–Margoliash, neighbour-joining and maximum likelihood methods. Although many of these trees (not shown) rearranged some of the mitochondrial DNA lineages in biologically unrealistic ways, they all grouped the 'dinosaur bone' sequences with the primate pseudogene sequences.

Such nuclear pseudogenes can complicate species phylogenies and population genetic studies that use mtDNA<sup>2</sup>, especially those that do not use protein-coding regions<sup>3,4,20</sup>. Fortunately, it is an easy matter to differentiate the anciently transferred primate cytochrome *b* pseudogenes from functional genes owing to frameshift mutations and stop codons (see also ref. 2). Recently transferred copies can be more problematical.

The facile amplification of primate nuclear pseudogenes using primers designed to match conserved regions of mammalian mtDNA has especially alarming implications for 'ancient' DNA studies. Because of the slower rate (Fig. 2) and random nature of nucleotide substitution in nuclear pseudogenes compared to mitochondrial genes, these transferred sequences can look like damaged 'molecular fossils', which is the expected appearance of authentic ancient DNA. Indeed, it has recently been proposed that the cytochrome *b*-like sequences amplified from 'dinosaur bone DNA' (using primers designed to conserved regions of mammalian mtDNA)<sup>6</sup> are due to human<sup>21</sup> nuclear DNA<sup>5</sup> contamination. Our analyses show that these 'dinosaur' sequences are most similar overall to the orangutan clade of anciently trans-

ferred primate cytochrome *b* pseudogenes (Table 1), and branch with them in evolutionary trees (Fig. 3). Furthermore, two of the 'dinosaur' sequences contain a single base insertion only found so far in the human nuclear pseudogene (Fig. 1)<sup>5</sup>. So, despite claims to the contrary<sup>22</sup>, the cytochrome *b* sequences amplified from 'dinosaur' DNA are almost certainly due to primate, most probably human<sup>5</sup>, nuclear pseudogene contamination. Moreover, the high level of nucleotide diversity among the nine 'dinosaur' sequences (Table 1) suggests that the contaminating DNA might have been damaged by the extensive decontamination procedures used in that study<sup>6</sup>.

Thus, contamination with nuclear pseudogenes<sup>3,5</sup> poses a serious hazard for forensic and ancient DNA studies. Independent laboratories might easily replicate the supposedly ancient sequences, which is one of the standards of proof for authentic ancient DNA<sup>21,23</sup>. In light of these new observations, other claims of ancient DNA sequences should be re-examined. When possible, nuclear DNA purified from humans, as well as other potential sources of contamination, should be included in the PCR controls. □

TABLE 1 Per cent nucleotide differences between the partial sequences of 'dinosaur' and other cytochrome *b* genes

|                  | 'Dinosaur bone DNAs' |    |    |    |    |    |    |    |    | Nuclear pseudogenes |    |    |    |    |       |    | Primate mtDNA |    |       |    |    |    |       | Bird mtDNA |    |       |    |    |    |    |
|------------------|----------------------|----|----|----|----|----|----|----|----|---------------------|----|----|----|----|-------|----|---------------|----|-------|----|----|----|-------|------------|----|-------|----|----|----|----|
| Sequence         | 15                   | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 8                   | 9  | 10 | 11 | 12 | 13    | 14 | 1             | 2  | 3     | 4  | 5  | 6  | 7     | 24         | 25 | 27    | 28 | 29 | 30 | 31 |
| Bone 4-37        | —                    | 19 | 19 | 16 | 16 | 22 | 20 | 25 | 20 | 9                   | 17 | 15 | 13 | 28 | 31    | 31 | 32            | 28 | 29    | 30 | 31 | 36 | 34    | 40         | 37 | 42    | 43 | 41 | 38 | 41 |
| Bone 31-44       | 19                   | —  | 13 | 14 | 11 | 17 | 21 | 26 | 22 | 24                  | 23 | 21 | 21 | 28 | 31    | 31 | 31            | 27 | 23    | 32 | 31 | 33 | 29    | 46         | 37 | 39    | 42 | 42 | 35 | 40 |
| Bone 2-61        | 19                   | 13 | —  | 14 | 9  | 18 | 21 | 24 | 24 | 26                  | 23 | 21 | 22 | 29 | 32    | 32 | 26            | 25 | 23    | 26 | 30 | 30 | 29    | 37         | 37 | 41    | 45 | 44 | 39 | 44 |
| Bone 2-18        | 16                   | 14 | 14 | —  | 12 | 17 | 17 | 26 | 23 | 21                  | 20 | 18 | 18 | 27 | 30    | 30 | 32            | 29 | 24    | 29 | 29 | 32 | 30    | 41         | 36 | 37    | 41 | 40 | 34 | 40 |
| Bone 20-61       | 16                   | 11 | 9  | 12 | —  | 17 | 20 | 27 | 22 | 24                  | 23 | 19 | 20 | 29 | 32    | 32 | 29            | 27 | 24    | 28 | 31 | 35 | 30    | 42         | 37 | 39    | 40 | 37 | 33 | 37 |
| Bone 2-37        | 22                   | 17 | 18 | 17 | 17 | —  | 20 | 33 | 27 | 29                  | 22 | 23 | 24 | 28 | 31    | 31 | 31            | 30 | 25    | 28 | 29 | 29 | 27    | 42         | 34 | 30    | 32 | 32 | 21 | 29 |
| Bone 3-37        | 20                   | 21 | 21 | 17 | 20 | 20 | —  | 24 | 23 | 22                  | 19 | 16 | 18 | 22 | 21    | 21 | 35            | 34 | 29    | 31 | 32 | 33 | 32    | 33         | 32 | 37    | 40 | 40 | 36 | 36 |
| Bone 5-37        | 25                   | 26 | 24 | 26 | 27 | 33 | 24 | —  | 17 | 28                  | 29 | 30 | 26 | 33 | 31    | 31 | 25            | 27 | 28    | 33 | 26 | 27 | 28    | 43         | 38 | 34    | 46 | 44 | 36 | 41 |
| Bone 6-37        | 20                   | 22 | 24 | 23 | 20 | 27 | 23 | 17 | —  | 23                  | 23 | 24 | 20 | 25 | 32    | 32 | 30            | 24 | 24    | 29 | 20 | 27 | 27    | 42         | 37 | 33    | 45 | 44 | 36 | 34 |
| Mean difference: | 19.9%                |    |    |    |    |    |    |    |    | 21.5%               |    |    |    |    | 29.1% |    |               |    | 28.9% |    |    |    | 38.3% |            |    | 38.2% |    |    |    |    |

Values shown in the matrix are the per cent nucleotide differences between the 'dinosaur bone' and other sequences shown in Fig. 1 (identified by number across the top line), corrected for missing data at the ends of some of the 'dinosaur' sequences, and rounded to the nearest whole number. The values shown are not corrected for multiple hits or transition bias. The average per cent differences between the 'dinosaur' and other sequences (bottom line) were calculated using the unrounded values.

Received 9 June; accepted 15 September 1995.

- Kocher, T. D. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6196–6200 (1989).
- Smith, M. F., Thomas, W. K. & Patton, J. L. *Molec. Biol. Evol.* **9**, 204–215 (1992).
- van der Kuyl, A. C., Kuiken, C. L., Dekker, J. T., Perizonius, W. R. K. & Goudsmit, J. *J. molec. Evol.* **40**, 652–657 (1995).
- Quinn, T. W. *Molec. Ecol.* **1**, 105–117 (1992).
- Zischler, H. et al. *Science* **268**, 1192–1193 (1995).
- Woodward, S. R., Weyand, N. J. & Bunnell, M. *Science* **266**, 1229–1232 (1994).
- Mullis, K. B. et al. *Cold Spring Harbor Symp. quant. Biol.* **51**, 263–273 (1986).
- Wrischnik, L. A. et al. *Nucleic Acids Res.* **15**, 529–542 (1987).
- Cohen, J. *Science* **268**, 22–23 (1995).
- Ausubel, F. M. et al. *Current Protocols in Molecular Biology* (Wiley, New York, 1989).
- Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. & Takahata, N. *Proc. natn. Acad. Sci. U.S.A.* **92**, 532–536 (1995).
- Li, W.-H., Ellsworth, D. L., Krushkal, J., Chang, B. H.-J. & Hewett-Emmett, D. *Molec. phylogenet. Evol.* (in the press).
- Saitou, N. & Ueda, S. *Molec. Biol. Evol.* **11**, 504–512 (1994).
- Brown, W. M., Prager, E. M., Wang, A. C. & Wilson, A. C. *J. molec. Evol.* **18**, 225–239 (1982).
- Lopez, J. V., Yuhki, N., Masuda, R., Modi, W. & O'Brien, S. J. *J. molec. Evol.* **39**, 174–190 (1994).
- Zullo, S., Sieu, L. C., Slightom, J. L., Hadler, H. I. & Eisenstadt, J. M. *J. molec. Biol.* **221**, 1223–1235 (1991).
- Shay, J. W. & Werbin, H. *Mutat. Res.* **275**, 227–235 (1992).
- Hu, G. & Thilly, W. G. *Gene* **147**, 197–204 (1994).
- Fukuda, M. et al. *J. molec. Biol.* **186**, 257–266 (1985).
- van der Kuyl, A. C., Kuiken, C. L., Dekker, J. T. & Goudsmit, J. *J. molec. Evol.* **40**, 173–180 (1995).
- Hedges, S. B. & Schweitzer, M. H. *Science* **268**, 1191–1192 (1995).
- Woodward, S. R. *Science* **268**, 1194 (1995).
- Handt, O., Höss, M., Krings, M. & Pääbo, S. *Experientia* **50**, 524–529 (1994).
- Anderson, S. et al. *Nature* **290**, 457–465 (1981).
- Gustincich, S., Manfioletti, G., Del Sal, G., Schneider, C. & Carninci, P. *BioTechniques* **11**, 298–302 (1991).
- Felsenstein, J. *PHYLIP (Phylogeny Inference Package) Version 3.5c* (Department of Genetics, University of Washington, Seattle, 1994).
- Swofford, D. L. *PAUP: Phylogenetic Analysis Using Parsimony Version 3.1* (Illinois Natural History Survey, Champaign, IL, 1993).
- Maddison, W. P. & Maddison, D. R. *MacClade: Analysis of Phylogeny and Character Evolution Version 3.04* (Sinauer, Sunderland, MA, 1994).
- Gingerich, P. D. *Molec. Biol. Evol.* **3**, 205–221 (1986).

ACKNOWLEDGEMENTS. We thank T. Disotell, B. Hedges, R. Higuchi, W. Messier, S. Pääbo, M. Ruvoilo and D. Shub for discussions; B. Hedges, A. van der Kuyl, W.-H. Li and S. Pääbo for preprints; T. Disotell, the late A. C. Wilson, Yerkes Regional Primate Center, the Houston Zoo, the Bronx Zoo, the St Louis Zoo, and Duke University Primate Center for primate specimens; D. Swofford and J. Felsenstein for computer programs. This work was supported by a grant to C.-B.S. from the NIH. C.-B.S. is an NSF Presidential Faculty Fellow.

## A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans

Hans Zischler, Helga Geisert, Arndt von Haeseler & Svante Pääbo

Zoologisches Institut der Universität München, PO Box 202136, D-80021 München, Germany

MAMMALIAN mitochondrial DNA sequences evolve more rapidly than nuclear sequences<sup>1</sup>. Although the rapid rate of evolution is an advantage for the study of closely related species and populations, it presents a problem in situations where related species, used as outgroups in phylogenetic analyses, have accumulated so much change that multiple substitutions obliterate the phylogenetic information<sup>2</sup>. However, mitochondrial DNA sequences are frequently inserted into the nuclear genome<sup>3</sup>, where they presumably evolve as nuclear pseudogene sequences and therefore more slowly than their mitochondrial counterparts. Such sequences thus represent molecular 'fossils' that could shed light on the evolution of the mitochondrial genome and could be used as outgroups in situations where no appropriate outgroup species exist. Here we show that human chromosome 11 carries a recent integration of the mitochondrial control region that can be used to gain further insight into the origin of the human mitochondrial gene pool.

Much of the debate about the origin of modern humans has focused on where the root falls in a phylogenetic tree relating the mitochondrial genomes of contemporary humans to each other<sup>2,4–8</sup>. Initial analyses of restriction maps of world-wide

samples of the mitochondrial genome<sup>4</sup> and DNA sequences of hypervariable parts of the mitochondrial control region<sup>5</sup> suggested that the root falls in Africa. However, subsequent re-analyses showed that there is very little statistical support for this conclusion<sup>6–8</sup>. A major limitation for any attempt to resolve this problem is the dearth of adequate outgroups. The control region of chimpanzees, probably the closest living relatives of humans<sup>9</sup>, is so divergent from the human counterpart that multiple substitutions make it suboptimal or even unsuited as an outgroup<sup>2</sup>. It is known that insertions of mitochondrial sequences into the nuclear genome occur (ref. 3 and refs therein), so we have searched the human nuclear genome for recent integrations of mitochondrial control region sequences in order to provide a more closely related outgroup to root the human mitochondrial (mt) DNA tree.

Human nuclear DNA was prepared from sperm heads<sup>10</sup>. An aliquot was restricted with *Bam*HI, which cuts human mitochondrial DNA once, and the digest was separated by electrophoresis and immobilized on a filter. When the filter was hybridized under stringent conditions with a probe containing the mitochondrial control region, it was found that a DNA preparation enriched for nuclear DNA showed the expected mitochondrial band of 16.5 kilobases(kb), plus a band of ~10 kb (Fig. 1a). In DNA isolated from sperm heads, the 10-kb band was still present although no mitochondrial DNA could be detected.

The nuclear mitochondrial insertion and its flanking sequences were isolated by a combination of low-stringency polymerase chain reaction (PCR) and anchor PCR, and sequenced (Fig. 1b). The integration comprises 540 base pairs (bp) (mtDNA positions 59 to 16,089 in ref. 11). Its ends coincide with a direct repeat of 6 bp in the mitochondrial sequence. Primers specific for the nuclear sequences flanking the insertion were constructed and used to amplify the insertion sequence from a set of human-rodent hybrid cells lines, each carrying one human chromosome. It could be shown that the insertion is located on human chromosome 11 (data not shown). The flanking primers were also used to isolate the site of integration from a chromosome lacking the insertion (see later). The non-insertion allele indicates that six bases, which exhibit four identities to the direct repeat at the ends of the mitochondrial sequence, were lost from the chromosomal sequence during the insertion event. It may be of significance that the ends of the insertion coincide approximately with the ends of the 7S DNA, the third DNA strand that makes up the triple-stranded D-loop portion of the mitochondrial control region. Thus, one of its ends is ~50 bp upstream of one 5' end of the 7S DNA<sup>12</sup>, whereas the other is 17 bp shorter than that of the 7S DNA<sup>13</sup>. An integration of the same portion of the control region has been described in the rat<sup>3</sup>. Thus, the 7S DNA may have a tendency to escape from mitochondria and integrate into the nuclear genome.

A total of 302 bp of the insert overlap with mitochondrial control-region sequences determined from 134 individuals from various human populations<sup>5</sup>. Table 1 shows the average numbers of transitions and transversions observed at these 302 positions among the control region sequences from the humans, among

TABLE 1 Average observed numbers of transitions and transversions among chimpanzee mtDNA, the insertion and human mtDNA sequences

|             | Chimpanzees | Insertion | Humans    |
|-------------|-------------|-----------|-----------|
| Chimpanzees | 15.5/3.5    | 35.8/17.4 | 31.1/17.1 |
| Insertion   |             |           | 17.9/ 2.9 |
| Humans      |             |           | 8.4/0.28  |

In the difference matrix, transitions are indicated by the first number and transversions by the second. For chimpanzee mtDNA sequences,  $n=60$ ; for human mtDNA sequences,  $n=134$ . Fifty-nine chimpanzee sequences were taken from ref. 18 and one from ref. 5; human data are from ref. 5, from which one sequence, which contained 70 unresolved bases in the alignment, was excluded.

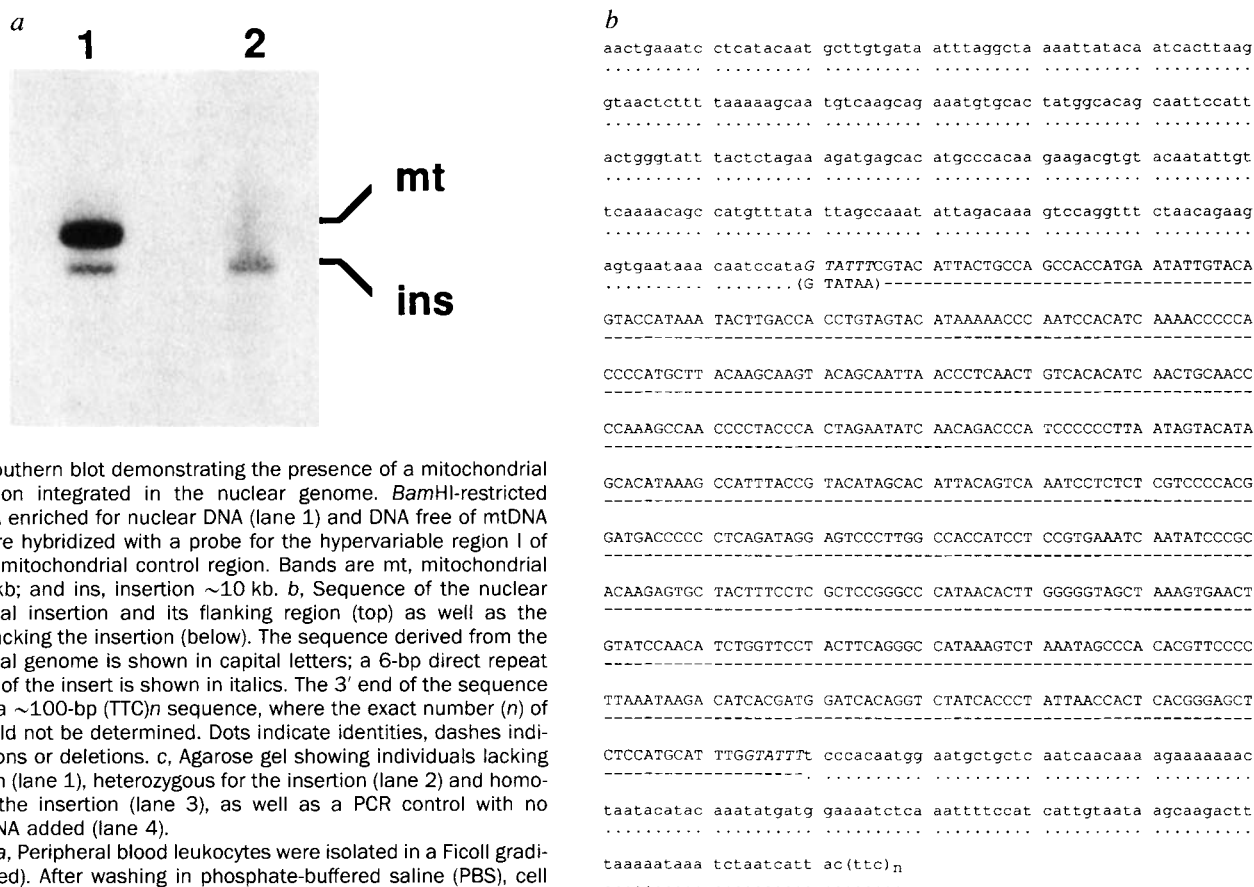
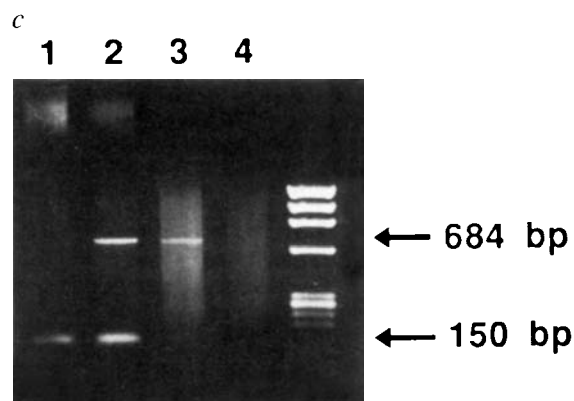


FIG. 1 a, Southern blot demonstrating the presence of a mitochondrial control region integrated in the nuclear genome. *Bam*HI-restricted human DNA enriched for nuclear DNA (lane 1) and DNA free of mtDNA (lane 2) were hybridized with a probe for the hypervariable region I of the human mitochondrial control region. Bands are mt, mitochondrial DNA, 16.5 kb; and ins, insertion ~10 kb. b, Sequence of the nuclear mitochondrial insertion and its flanking region (top) as well as the sequence lacking the insertion (below). The sequence derived from the mitochondrial genome is shown in capital letters; a 6-bp direct repeat at the ends of the insert is shown in italics. The 3' end of the sequence consists of a ~100-bp (TTC)<sub>n</sub> sequence, where the exact number (n) of repeats could not be determined. Dots indicate identities, dashes indicate insertions or deletions. c, Agarose gel showing individuals lacking the insertion (lane 1), heterozygous for the insertion (lane 2) and homozygous for the insertion (lane 3), as well as a PCR control with no template DNA added (lane 4).

**METHODS.** a, Peripheral blood leukocytes were isolated in a Ficoll gradient (Nycomed). After washing in phosphate-buffered saline (PBS), cell membranes were lysed for 20 min in the presence of 1.2% NP-40 in PBS at 37 °C (lane 1) and nuclei were pelleted (800g, 5 min, 4 °C). This procedure was repeated twice before DNA was isolated from the nuclei. A semen sample (lane 2) from the blood donor was centrifuged (800g, 5 min, 4 °C) and the sperm cells were washed twice in 50 ml PBS. Differential lysis of sperm cells was done as described<sup>10</sup> except that lysis in the absence of DTT was repeated 5 times between pelleting of the sperm nuclei. DNA isolation, restriction, blotting and probing were done according to standard protocols<sup>19</sup> using a radioactively labelled probe generated by PCR using the primers L16087 (5'-CACCCATCAACA-ACCGCTAT-3') and H16401 (5'-TGATTTACGAGGAGGATGGT-3') (ref. 11). Washes were done in 0.1 × SSC/0.1% SDS at 68 °C. The filter was exposed for 48 h at -80 °C with intensifying screens. Additional restriction analysis confirmed that the 10-kb band was not localized in the mitochondrial genome (not shown). b, 100 µg of mitochondrial-free nuclear DNA was digested with *Pvu*II and separated by electrophoresis in an agarose gel. DNA of ~3.5 kb was isolated from the gel according to standard protocols<sup>19</sup>. On this template, the 5'-biotinylated primer H16438 (5'-CCCCGAGCGAGGAGAGTAGC-3') was extended during 50 cycles with *Taq* polymerase (1 min annealing at 68 °C; 1.5 min extension at 72 °C; 40 s denaturation at 92 °C). Excess primers were removed by ultrafiltration and extension products were isolated by adsorption to magnetic beads (from Dynal A/S). A homo-G-tail was added using terminal transferase (from NEB) according to the supplier's protocol. Unbiotinylated H16438 and a C<sub>15</sub>-mer were used to amplify the extension products by PCR (1 min annealing at 60 °C; 1.5 min extension at 72 °C; 40 s denaturation at 92 °C). After cloning (pCR-Script from Stratagene) in *E. coli* (SURE strain, Stratagene), colonies were screened with the probe for hypervariable region I. Positive plasmids were isolated, restricted, and the longest inserts sequenced. For the second flank, size-enriched DNA served as template for the primer 5'-ACAAAGTCCAGGTTTCTAACAG-3', which flanks the insert, in a low-stringency single



primer (LSSP) PCR as described<sup>20</sup>, except that annealing was at 35 °C and 35 cycles were performed. Aliquots of the LSSP-PCR were subjected to a semi-nested second PCR with this primer and the primer 5'-CCAG-GTTTCTAACAGAAGAGTG-3'. Products were cloned and the largest inserts sequenced. c, Primers flanking the insertion (5'-ACAAAGTCCAG-GTTTCTAACAG-3' and 5'-AGTCTTGCTTATTACATGATGG-3') were used to amplify the insertion locus from 300 ng genomic DNA. 35 cycles of PCR consisted of annealing at 63 °C for 1 min; extension at 72 °C for 1 min; denaturation at 92 °C for 40 s. PCR products were resolved on a 2.5% agarose gel (Seakem) and visualized by ethidium bromide staining. When four common chimpanzees, one bonobo, and four gorillas were tested with these as well as primers internal to the insert, they did not show any evidence of an insertion.

60 chimpanzees, as well as between the human and chimpanzee sequences, and the insertion. The insertion differs by an average of 53 (range 48–62) substitutions from the chimpanzee sequences and by an average of 21 (range 17–25) substitutions from the human sequences. Thus, the insertion displays substantially more sequence similarity to the human than to the chimpanzee

mitochondrial sequences. Furthermore, using primers flanking the insertion, we did not detect it in chimpanzees or gorillas. We conclude that the insertion took place after the divergence between humans and chimpanzees.

Table 1 also shows that the human sequences differ by an average of 8.7 (range 0–19) substitutions among each other,



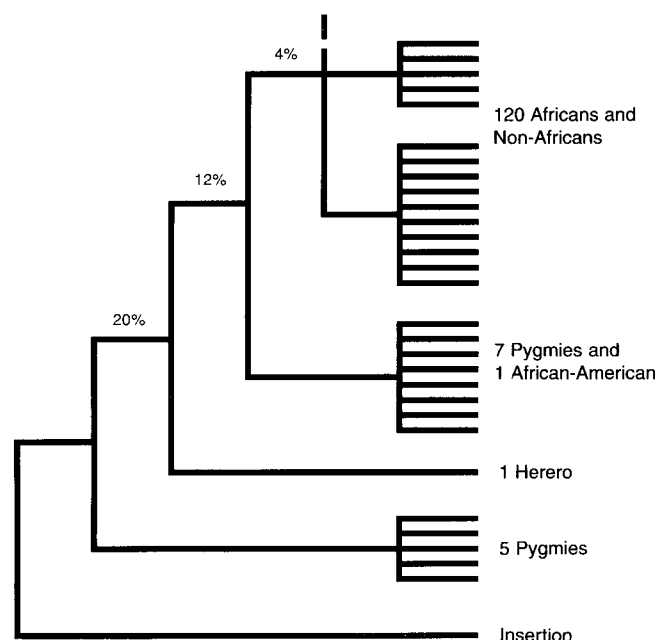


FIG. 2 Neighbour-joining consensus tree of 134 human mitochondrial control region sequences constructed using the insertion as an outgroup. For clarity, only the first five branches and their geographical states are indicated; their internal branching structures are not shown. The three most basal branches consist of the sequences: 41–45 (first branch); 34 (second branch); 33, 37–40, 46–48 (third branch) (numbering as in ref. 5). Broken line (top) points to the remaining branches of the tree, included in the 120 African and Non-African sequences not present in the first three branches. Numbers on internal branches refer to neighbour-joining bootstrap probabilities.

**METHODS.** The consensus tree shown was computed from 100 bootstrap replications using the neighbour-joining algorithm and maximum likelihood distances as implemented in *Phylip*, version 3.5p (ref. 21). As the neighbour-joining algorithm may be input-order dependent, the input order was randomized 1,000 times and a consensus tree constructed. This tree contained the same three basal branches, made up of the same sequences, as the tree shown. To test further whether the insertion represents an outgroup to the human mitochondrial gene pool, trees with the 134 human sequences, the insertion and 60 chimpanzee sequences were constructed. Both neighbour-joining ( $P=0.29$ ) and parsimony trees showed the insert to be a sister taxon to the human sequences. Further tree analyses of the 26 human sequences (1–6, 16, 17, 32, 41–45, 58, 74, 75, 81, 84, 87, 88, 91, 95, 102, 105, 110; numbered as in ref. 5), that had pairwise sequence differences equal to or larger than the smallest distance between any human sequence and the insertion, revealed that, whereas parsimony failed to show the insertion as an outgroup to this subset of human sequences, it falls as the sister group of the human sequences in the neighbour-joining tree ( $P=0.45$ ).

whereas they differ by 21 (range 17–25) substitutions from the insertion sequence. Thus a small proportion (0.65%) of the pairwise comparisons between human mitochondrial sequences show the same number or more differences than some comparisons between mitochondrial sequences and the insertion. Neighbour-joining and parsimony trees were constructed for the chimpanzee sequences, the insertion and the human sequences. These trees show the insertion to be a sister taxon of the current human sequences. Thus, although the phylogenetic analyses might leave some room for doubt (Fig. 2 legend), the most likely time that the common ancestor of the insertion sequence and the contemporary mitochondrial gene pool existed is shortly before the existence of a common ancestor of the current mitochondrial gene pool.

Primers specific for nuclear DNA sequences flanking the insertion were used to screen populations around the world for the presence of the insertion. Figure 1c shows that individuals heterozygous as well as homozygous for the insertion could be detected. Overall, 39% of chromosomes tested carried the insertion. In four African populations, the frequency of chromosomes carrying the insertion ranges between 10 and 25%, whereas it varies between 38% and 78% in populations tested in Europe, Asia, Oceania and South America (Table 2). As the amount of

mitochondrial sequence diversity is larger in Africa than outside Africa (see for example, refs 4, 14), the most likely explanation for this difference in frequency is that the populations outside Africa have experienced a bottleneck, which presumably was associated with the colonization of Eurasia from Africa.

To determine whether nucleotide changes have occurred in the insertion since its transfer to the nuclear genome, five insertions were sequenced from different geographical regions (three from Europe, one each from southern Africa and Japan). All sequences were identical (results not shown). As the insertion shows little or no sequence polymorphism and has not yet reached fixation in the human population, it has probably not accumulated any substantial changes since its transfer to the nuclear genome. This is consistent with the rate of evolution of non-coding sequences in the nuclear genome<sup>15</sup> being of the order of tenfold<sup>16</sup> to fortyfold<sup>17</sup> slower than that of the mitochondrial control region. Thus, the insertion represents a molecular 'fossil' which, with very few or no changes, has been transmitted in the human germ line since its transfer into the nuclear genome. This makes it well suited to be a potential outgroup for a tree of mtDNA sequences of modern humans—not only because it is more closely related to modern human than to chimpanzee sequences and is thus less affected by multiple substitutions, but also because it has undergone very little or no change since integration.

Mitochondrial control-region sequences from 134 humans were used previously in attempts to reconstruct the origin of the human mitochondrial gene pool<sup>5</sup>. Here we used them to construct a phylogenetic tree by the neighbour-joining algorithm, with the insertion sequence rather than the chimpanzee sequences as outgroup. Figure 2 shows that this tree contains one basal branch leading to exclusively Pygmy sequences, followed by two branches leading to Herero, Pygmy and African American sequences. In the fourth branch from the root, the first non-African sequences occur. Thus the topology of this tree supports an African origin of the human mitochondrial gene pool. Although the statistical support for the basal 'African' branches is weak ( $P=0.20$ , 0.12 and 0.04, respectively), it is greater than in the case in which a chimpanzee was used as outgroup<sup>7</sup>. Strong support for the topology should not be expected when the number of taxa used is much larger than the

TABLE 2 Frequencies of insertion in 12 populations

|          | Population    | Insertion frequency (%) | Chromosomes sampled |
|----------|---------------|-------------------------|---------------------|
| Africa   | Mbuti         | 10                      | 50                  |
|          | Biaka         | 20                      | 30                  |
|          | San           | 10                      | 20                  |
|          | Bantu         | 25                      | 20                  |
| Eurasia  | Europeans     | 54                      | 26                  |
|          | Yemenite Jews | 43                      | 14                  |
|          | Thoti         | 38                      | 26                  |
|          | Japanese      | 65                      | 20                  |
| Americas | Surui         | 78                      | 18                  |
|          | Quechua       | 54                      | 26                  |
| Oceania  | Melanesians   | 68                      | 22                  |
|          | Goroka        | 50                      | 20                  |

number of variable positions. But it is noteworthy that when a data set of 186 African, 534 European and 93 Native American sequences (but excluding the sequences in ref. 5) are used, the root of the tree again falls in Africa (E. Watson, A. Sajantilla, R. Ward and H. Z., data not shown). Thus, when a previously unavailable outgroup is used, either with the original or with other data sets, the results again point to an African origin of the human mitochondrial gene pool.

In conclusion, the isolation of nuclear 'fossils' of mtDNA represents a means to reconstruct the evolution of the mitochondrial genome as well as a way to obtain outgroups for phylogenetic analyses where suitable extant taxa are not available. We expect such sequences to increase the quality of information gained in studies of mtDNA sequence variation. □

Received 3 April; accepted 6 October 1995.

1. Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. *J. molec. Evol.* **18**, 225–239 (1982).
2. Maddison, D. R., Ruvo, M. & Swofford, D. L. *Syst. Biol.* **41**, 111–124 (1992).
3. Zullo, S., Sieu, L. L., Sligham, J. L., Hadler, H. I. & Eisenstadt, J. M. *J. molec. Biol.* **221**, 1223–1235 (1991).
4. Cann, R. L., Stoneking, M. & Wilson, A. C. *Nature* **325**, 31–36 (1987).
5. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. *Science* **253**, 1503–1507 (1991).
6. Templeton, A. R. *Science* **255**, 737 (1992).
7. Hedges, S. B., Kumar, S., Tamura, K. & Stoneking, M. *Science* **255**, 737–739 (1992).
8. Maddison, D. R. *Syst. Zool.* **40**, 355–363 (1991).
9. Ruvo, M., Disotell, T. R., Allard, M. W., Brown, W. M. & Honeycutt, R. L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1570–1574 (1991).
10. Gill, P., Jeffreys, A. J. & Werrett, D. J. *Nature* **318**, 577–579 (1985).
11. Anderson, S. et al. *Nature* **290**, 457–465 (1981).
12. Tapper, D. P. & Clayton, D. A. *J. biol. Chem.* **256**, 5106–5115 (1981).
13. Doda, J. N., Wright, C. T. & Clayton, D. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6116–6120 (1981).
14. Pult, I. et al. *Biol. Chem. Hoppe-Seyler* **375**, 837–840 (1994).
15. Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York, 1987).
16. Ward, R. H., Frazier, B. L., Dew-Jager, K. & Pääbo, S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8720–8724 (1991).
17. Lundstrom, R., Tavaré, S. & Ward, R. H. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5961–5965 (1992).
18. Morin, P. A., Moore, J. J., Chakraborty, R., Jin, L., Goodall, J. & Woodruff, D. S. *Science* **265**, 1193–1201 (1994).
19. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
20. Pena, S. D. J. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 1946–1949 (1994).
21. Feisenstein, J. *Phylip* version 3.5p (Univ. Washington, Seattle, 1994).

**ACKNOWLEDGEMENTS.** We thank K. and J. Kidd, S. Tishkoff, T. Jenkins, M. Batzer, B. Bonne-Tamir and W. Schempp for DNA samples; A. Sajantilla, R. H. Ward and E. Watson for sharing unpublished data; V. Börner, M. Höss, M. Krings and H. Poinar for discussion; S. Tisborn for contributions in the early phase of this project; W. Schartau for oligonucleotide synthesis; and the DFG for financial support.

## Visual cortical mechanisms detecting focal orientation discontinuities

Adam M. Sillito, Kenneth L. Grieve,  
Helen E. Jones, Javier Cudeiro\* & Justin Davis

Department of Visual Science, Institute of Ophthalmology,  
Bath Street, London EC1V 9EL, UK

\* Departamento de Ciencias de la Salud I, (Univ. La Coruna),  
Unidad de Cirugía Experimental Hosp. Juan Canalejo,  
As Xubias 84, Spain

**NEURONS in the primary visual cortex (V1) respond in well defined ways to stimuli within their classical receptive field, but these responses can be modified by stimuli overlying the surrounding area<sup>1–7</sup>. For example patch-suppressed cells respond to gratings of a specific orientation within their classical receptive field, but the response diminishes if the grating is expanded to cover the surrounding area<sup>1–7</sup>. We report here more complex effects in many such cells. When stimulated at their optimal orientation, introducing a surrounding field at a significantly different (for example, orthogonal) orientation enhanced their output by both a disinhibitory mechanism and an active facilitatory mechanism producing**

**'supra-optimal' responses. Importantly, some cells responded well if the orientations of centre and surround stimuli were swapped. The output reflected the discontinuity because neither stimulus component alone was effective. Under these stimulus conditions simultaneously recorded cells with orthogonally oriented receptive fields showed correlated firing consistent with neuronal binding to the configuration. We propose a mechanism integrating orientation-dependent information over adjacent areas of visual space to represent focal orientation discontinuities such as junctions or corners.**

The data were drawn from recordings of simple and complex cells with arrays of three electrodes inserted in the primary visual cortex of cat and macaque monkey. We centred a circular patch of a sinusoidal grating stimulus on the receptive field of one of the cells and recorded its response to drifting of the grating within this circular window while varying the orientation and diameter of the window. This procedure was repeated with an annulus of grating, also centred over the field, but with the diameter of its inner wall varying over the same range as the central window (Fig. 1). In the monkey the majority of cells (33/36) studied showed response suppression (mean 68%) with increasing grating patch diameter as previously reported<sup>7</sup>. An example is shown in Fig. 1a. The area of each rectangle (in the matrix) corresponds to the firing rate elicited by that particular combination of orientation and patch diameter. In this example it appeared that as the stimulus diameter increased beyond 1° the stimulus encroached on an inhibitory zone in the receptive field which reduced the response. The situation is, however, more complex; for example, at the optimal orientation either a 2° patch or an annulus with an inner diameter of 2° exerted an excitatory effect, but the two stimuli were ineffective when combined (Fig. 1b). Moreover, the response to the inner/outer combination is dependent on their relative orientations (Fig. 1c). The orientation of the inner patch varies along the x axis and the outer along the y axis. When either the inner or the outer stimulus was at an orientation in the range –10° to +45° (that is, around its preferred orientation) the cell responded provided the other stimulus was at a markedly different orientation. A further example is shown in Fig. 1d, where with a small window size (0.75°) both inner and outer stimuli drove the cell at all orientations, except when they were both at the same orientation. Thus the output of these cells seemed to signal an orientation discontinuity over the field rather than orientation *per se*.

The extent to which the outer stimulus drove cells varied, but even when it had no excitatory effect alone, it could still exert a profound influence on the response to the inner stimulus (Fig. 2a). The orientation tuning curves show the cell's response to the inner stimulus alone, the outer stimulus alone and to the presence of both with the inner held at the cell's optimal orientation while varying the orientation of the outer. The diameter here was 1° and corresponded to the diameter of the inner alone that elicited a maximal response from the cell. The outer stimulus alone produced no response but when paired with the inner it caused a huge potentiation of the response when its orientation deviated from the cell's optimal orientation. This effect was still seen when the diameter was increased from 1° to 3° (Fig. 2b), although the absolute magnitude of the response was decreased. We tested this cell with diameters from 0.5–5.0° and the largest response was seen to the cross oriented bipartite stimulus at 1°. There are thus two issues, the cross-orientation potentiating effect of the outer stimulus and the presence of a 'spatial focus' where the maximal response was elicited. The interpretation of the processes underlying these effects is complex because, as Fig. 1 shows, the inhibitory and excitatory mechanisms driven by the inner and outer stimuli seem to overlap. In particular the inhibitory mechanism of the 'surround suppression' at iso-orientation conditions seems to be enabled only when both centre and outer regions of the receptive field are stimulated. Both are of course stimulated when larger patch sizes are used. This type of effect has been previously reported for the cat visual cortex<sup>8–10</sup>.

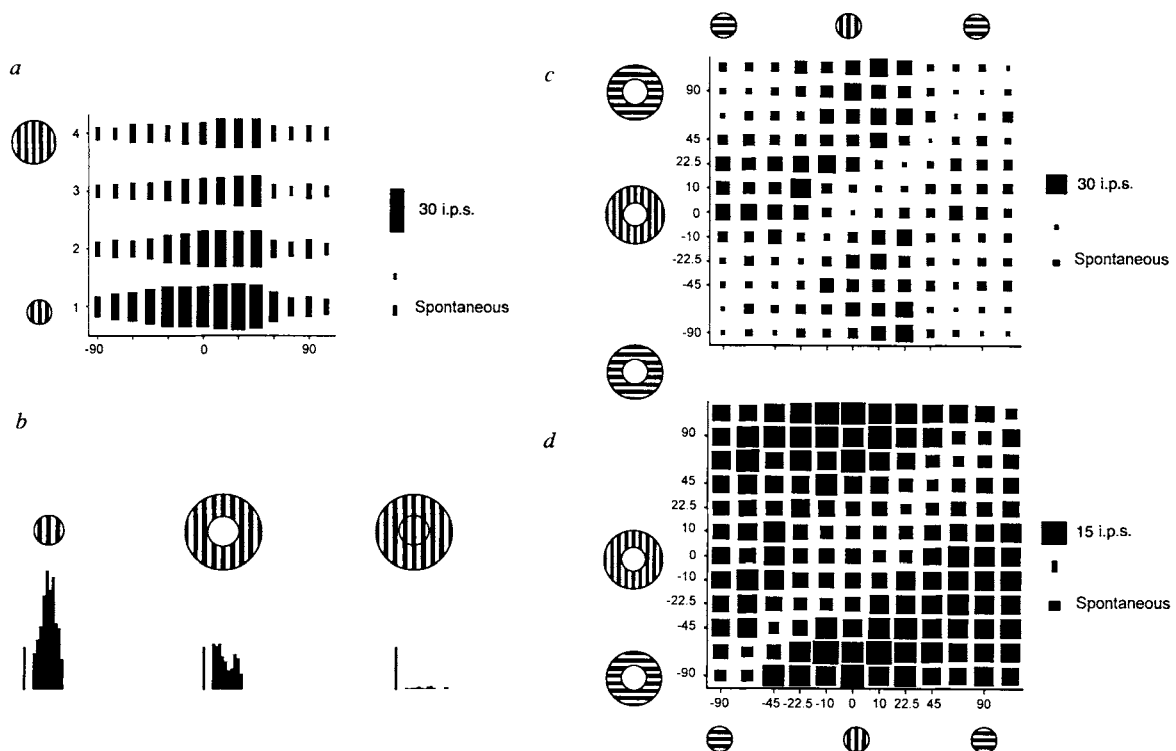


FIG. 1 Spatial interactions in the receptive field of cells in primary visual cortex. These records illustrate how a surround stimulus can modify the response of cells to a central stimulus over their receptive field. *a*, Response of a simple cell in primate V1 to a single grating patch. A patch of drifting sinusoidal grating (grating drifting within the circular patch) was centred on the receptive field. The orientation (x axis) and diameter (y axis) of this patch were varied and the magnitude of the cell's response (computed as the mean response rate in impulses per second (i.p.s.) averaged from the full stimulus cycle for 50 stimulus repeats) is represented by the area of the rectangle. The orientation is shown as a variation around 0°, where 0° was set to be close to the cell's optimal orientation. Stimuli were presented in a randomized interleaved sequence. The cell's response was sensitive to both the orientation and diameter of the stimulus, decreasing markedly at the larger diameters. Calibration sizes are given on the right of the figure and show the spontaneous level and the range from 1 i.p.s. to the value indicated. Grating spatial frequency 2 cycles per degree (c.p.d.), temporal frequency 4 Hz, contrast 0.36. See text for details. *b*, Peristimulus histograms (PSTHs) showing the response of the simple cell to a 2° patch alone, an annulus of grating with a 2° diameter inner wall and both together. All stimuli at the cell's optimal orientation. Both the patch and the annulus produce an excitatory response but their concatenation does not. Records are averaged from 50 stimulus repeats, with 25 ms bins. Mean response calculated as for *a*. Vertical calibration shows range for 0–40 i.p.s. Other details as for *a*. *c*, This extends the observations in *b* and shows the interaction between the patch and annulus for the 2° condition when their relative orientations are varied (randomized

interleaved sequence). In each case the orientation was varied around 0° which was set to be close to the cell's optimal orientation. Patch orientation is shown on the x axis, annulus orientation on the y axis. Area of the squares corresponds to response magnitude, details as for *a*. Each square shows the response for a particular combination of inner and outer orientations. Both the inner and outer stimuli will drive the cell for values around the cell's optimal orientation, but when their orientation is the same there is no response. Other details as for *a*. *d*, Further example of spatial interactions in an S-cell field with the two-component stimulus. The diameter of the central patch and inner annulus is 0.75° in this case. At this diameter the inner and outer stimulus components exert an effect at all orientations, except when they are both at the same orientation (compare with *b*) then there is a clear reduction in cell activity at all orientations (diagonal from bottom left to top right). Data averaged from 25 stimulus repeats. 2 c.p.d., 3 Hz, 0.36 contrast. Other details as for *a*. Note that this was an orientation-tuned cell, as the diameter of the window increased the responses fell into an orientation-linked pattern but with the iso-orientation deletions as in *c*. All these observations were made in the visual cortex of anaesthetized (sufentanil 4  $\mu\text{g kg}^{-1} \text{h}^{-1}$  or halothane (0.1–0.4%) in 70%  $\text{N}_2\text{O}/30\% \text{O}_2$ ) and paralysed (0.1  $\text{mg kg}^{-1} \text{h}^{-1}$  vecuronium bromide or 10  $\text{mg kg}^{-1} \text{h}^{-1}$  gallamine triethiodide) primates (macaque) or cats using methods described in detail elsewhere<sup>24</sup>. Before taking data detailed plots were made of classical receptive fields and boundaries and receptive field centre checked quantitatively with stationary flashing and moving stimuli.

The question here is whether the presence of the outer stimulus at non-optimal orientations is actually disinhibiting the cell from an input driven by the inner stimulus and thus releasing more of the direct excitatory drive of the inner, or bringing an additional cross-orientation facilitatory influence to bear on the cell. Our data suggest the latter; for example, Fig. 3*a* shows the response of a cell when varying the orientation of the inner patch alone, and that in 3*b* the response when varying the outer alone. The cell responded to the inner at orientations roughly  $\pm 20^\circ$  either side of its optimal. It gave no response at  $+90^\circ$  or  $-90^\circ$ . The outer alone failed to elicit a response at any orientation. However, with the inner at  $-90^\circ$  and the outer at orientations in the range  $0-10^\circ$  (Fig. 3*c*) the cell responded strongly, showing the best response for the outer at  $10^\circ$ . Expressed another way,

the presence of the outer stimulus at  $10^\circ$  enabled the cell to respond to the inner at  $-90^\circ$ , the combination rather than the components being effective. Similar cross orientation 'facilitatory' effects were seen in cat V1 (Fig. 3*d-f*). Like Fig. 3*a-c* this shows the response of the cell to the inner stimulus alone, outer alone and then the inner at  $90^\circ$  with the outer varying. The key point is that as the outer reached the cell's optimal it caused the cell to respond vigorously. Again the outer seemed to enable the cell to respond to the inner at  $90^\circ$  to its optimal, but this may be a misleading view, as the cell was responding to the presence of both at this configuration, neither alone could be considered to drive it. The response was an 'emergent' property of the two stimuli together. These cross-orientation effects were seen in over half of the cells in V1 that showed alignment inhibition (primate

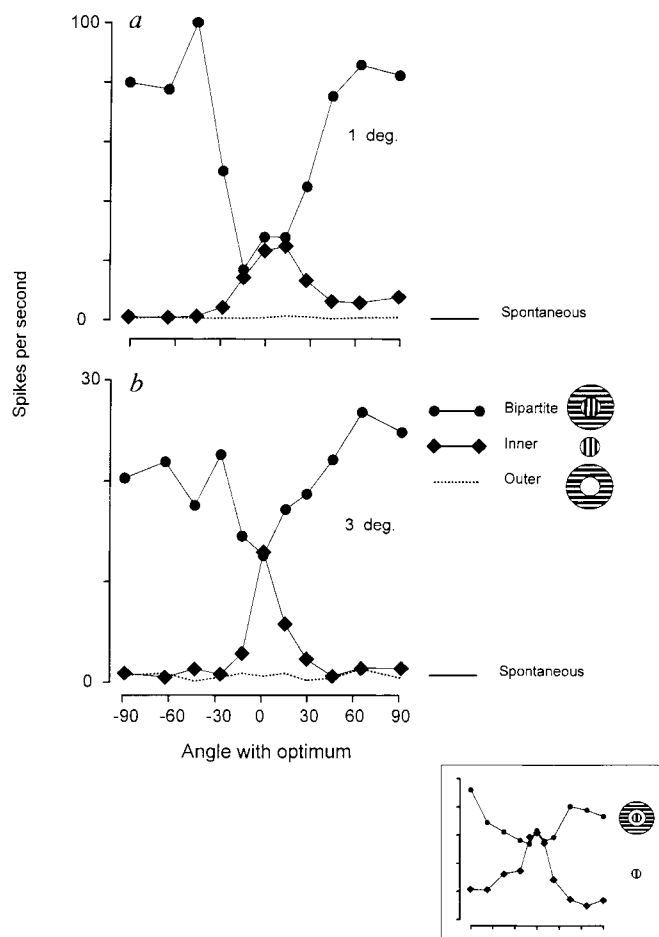


FIG. 2 Records showing cross-oriented stimulus components causing supra-optimal response in a simple cell in primate V1. This cell was strongly patch-suppressed. It gave its maximal response to the central patch alone at a diameter of  $1^\circ$  and the response decreased as patch diameter increased beyond this value. These curves show the mean response of the cell to the inner stimulus alone, the outer stimulus alone and both components together with the inner held at the cell's optimal orientation while the outer varied. *a*, This refers to data for the  $1^\circ$  patch combination. The annulus alone has no effect on the cell. However, as the orientation of the annulus is varied with the inner patch present at its optimal orientation, the outer stimulus strongly modifies the cell's response. For orientations of the outer between  $30^\circ$  and  $90^\circ$  from the cell's optimal the output of the cell is hugely increased. Thus the configuration of the inner patch at the optimal orientation and the outer at an orientation differing from the optimal by more than  $30^\circ$  appeared to constitute the 'optimal' stimulus for this cell. *b*, This shows the same protocol as *a* but the window for the inner patch/annulus combination is  $3^\circ$ . There is a marked reduction in response magnitude here as is reflected by the change in the maximum mean response from roughly 100 i.p.s. to less than 30. The presence of the 'cross-oriented' outer still results in an increase in response magnitude but it is much smaller. These data were derived from the mean response to 30 stimulus repeats as detailed in *a*, stimulus was 2 c.p.d., 3 Hz, contrast 0.36. As a control for the possible influence of the higher-order spatial transforms linked to the junction of the inner and outer stimuli, the curves in the inset square show that introducing a buffer zone between inner and outer stimuli does not eliminate the pattern of effect.

FIG. 3 Emergent excitatory responses to cross-oriented stimulus configurations in primate (*a-c*) and cat (*d-f*) V1 cells. *a*, Primate simple cell. Orientation tuning curve for  $0.75^\circ$  patch. *b*, Orientation tuning curve for an annulus with  $0.75^\circ$  inner diameter: orientation has no effect. *c*, Both components together with the inner patch at  $-90^\circ$  and the outer varying. The inner component alone at  $-90^\circ$  exerted no effect (see arrow) and the outer alone no effect at any orientation, but the combination of the two produces a strong response when the outer orientation is in the range  $0-10^\circ$ . The arrows in *a* and *b* identify the responses to the components, the arrow in *c* shows the response to the combination. The cell responded to the combination but not the individual components. Response is mean of 25 stimulus presentations as detailed in Fig. 1*a*, stimulus was 3 c.p.d., 3 Hz, 0.36 contrast. *d*, Cat special complex cell. Orientation tuning curve for inner alone with  $3^\circ$  patch. *e*, Tuning curve for outer alone with  $5^\circ$  inner border diameter, there is no response to this. *f*, Interaction of both components with a  $5^\circ$  partition of the components, the inner is held at  $90^\circ$  (see arrow in *d*) and the outer is varied. As the outer reaches the cell's optimum the cell gives a strong response despite the fact that neither component alone was effective for this combination of orientations (see arrows in *d* and *e*). The cell responded to the configuration not the elements. Data taken from 40 stimulus presentations, mean response computed as in Fig. 1*a*, stimulus was 0.66 c.p.d., 2 Hz, 0.36 contrast.

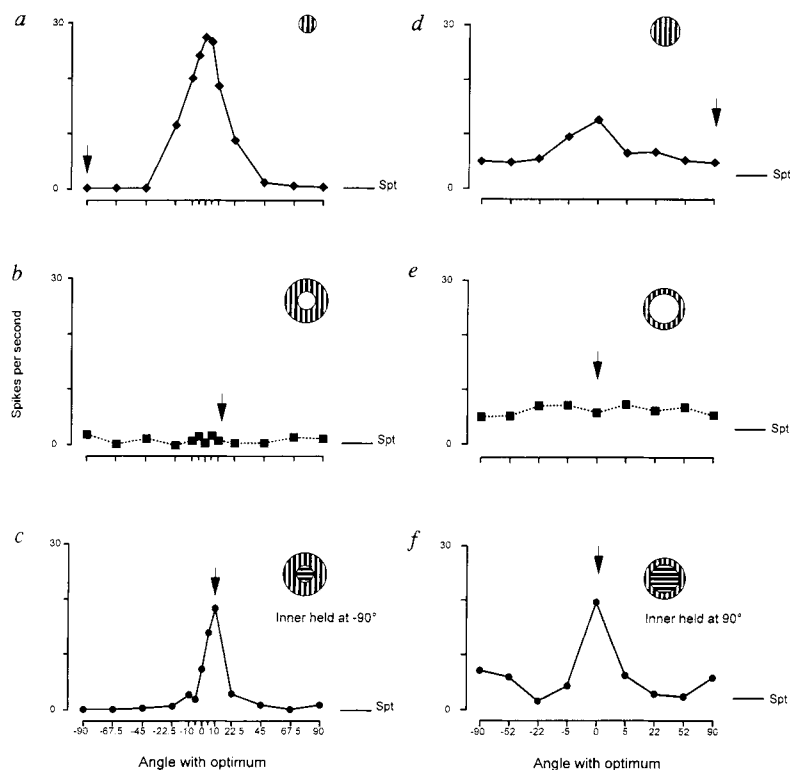
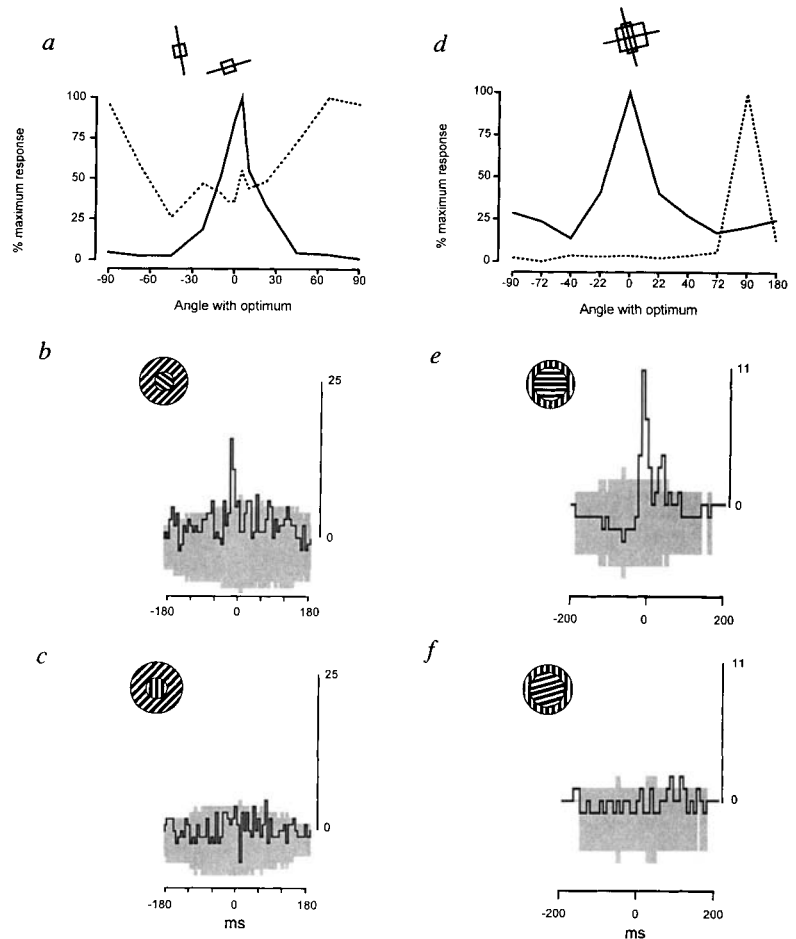


FIG. 4 Evidence of functional connectivity between pairs of simple cells with cross-oriented fields in primate and cat V1. Because we recorded from several cells simultaneously at each location with our triadic electrode arrays we obtained data from the whole group for tests focused on any one of the cells recorded. Thus when we were running the detailed tests shown in Figs 1–3 on the particular cells illustrated we had simultaneous recordings of the responses of the other cells to these stimuli. Some of these cells had overlapping or nearly overlapping fields and others had non-overlapping fields. Similarly some had the same and some had different orientations. For any given cell we could thus check how the other cells were responding to the stimulus parameters used to test its responses. We obtained data from 37 pairs of cells of which 14 pairs had cross-oriented or near cross-oriented receptive fields. For these pairs, when the cell under review was being tested we could check how the varying stimulus configuration influenced the way its activity correlated with that of the other cell (to the same stimulus). For the cross-oriented pairs we found correlations when a bipartite cross-oriented stimulus was driving the cell under review and producing the effects shown in Figs 2 and 3. The two illustrations here come from primate (a–c) and cat (d–f) and comprise pairs with non-overlapping (a–c) and overlapping (d–f) receptive fields as examples. a, d, Show the receptive field locations and the orientation tuning curves of the cells. Solid curves show orientation tuning of reference cell, interrupted curves that of the other cell (these were determined for each cell with a patch of grating of varying orientation centred over their receptive field). b, c, e, f, Show the cross-correlograms obtained from the two spike trains of the cell pairs when driven by the stimulus configurations illustrated to the top left of each correlogram. Synchronized firing, indicative of connectivity, shows as a central peak in the cross-correlogram, points outside the shaded area exceed the 99.5% confidence limits<sup>12–14</sup> for significant effects, so peaks in b and e for cross-oriented inner and outer stimuli are highly significant. Varying the relative orientation of the inner and outer sections of the stimulus eliminates the peaks indicating that the synchronized firing is dependent on stimulus configuration. Cross-correlograms were corrected for effects that would simply follow from the overlapping discharges caused by the coordinated motion of



the stimulus components<sup>12–14</sup>. Bin size for cross-correlograms, 8 ms and vertical calibration indicates counts per bin. Symbols show bipartite stimulus used, symbols not to scale of receptive field maps.

17/28, cat 8/13). The remainder showed the iso-orientation inhibition between the inner and outer stimuli, but for any given inner stimulus diameter did not show an increment in response magnitude for the cross-orientation difference between inner and outer.

These results indicate a facilitatory interaction between elements of V1 circuitry representing markedly different orientations in contradiction to the common belief that functional connectivity is only seen between cells of like orientation<sup>11</sup>. To obtain further evidence for functional connectivity between cells with cross-oriented fields, we looked for synchronized firing of pairs of cells. We obtained data from 37 pairs of cells in primate (17/37) and cat (20/37) V1, of which 14 pairs (8 primate, 6 cat) had cross-oriented or near cross-oriented receptive fields (range 67.5–100°). All the cross-oriented cell pairs, when stimulated with cross-oriented bipartite gratings, showed clear peaks of synchronized firing (>99.5% confidence limit<sup>12–14</sup>) in the cross-correlogram, suggesting functional connectivity<sup>15–17</sup>. Figure 4a–f shows the receptive fields, orientation tuning curves and cross-correlation functions for pairs of cross-oriented cells in primate and cat V1. These showed synchronized firing to a cross-oriented bipartite stimulus as shown in Fig. 4b, e with highly significant peaks (>99.5%) in the cross-correlograms. These were lost when the relative orientation of the two components was shifted (Fig. 4c, f). The connectivity of the 'network' thus seems to be dynamically organized to enable the representation of the discontinuity by allowing links between cells with cross-oriented fields.

These observations suggest the presence of a neural organization in V1 that may represent a mechanism for detecting a

local orientation discontinuity<sup>18–20</sup>. The receptive field of the cells involved needs to be conceived in terms of the discontinuity rather than any arrangement that can be mapped with a single stimulus. Cells that show this behaviour include some but not all of those that would be classically considered hypercomplex or end-stopped<sup>21</sup>; they also include some that, although patch-suppressed, would not be considered end-stopped<sup>7,22</sup>. It has long been recognized that hypercomplex cells in particular are inhibited by surrounding stimuli at their optimal orientation<sup>9,10,22</sup> and that their response is less attenuated if the surrounding stimulus deviates from the optimal orientation. It has also been suggested that this effect may be linked to perceptual pop-out<sup>23</sup>. The present results reveal a much more complex and focal process; first, the cells' responses seem strongly suppressed at all orientations (not just the optimal) for an iso-orientation stimulus (see Fig. 1c, d). Second, where there is an orientation discontinuity there appears to be both a loss of the suppression and an active enhancement of responses. Finally, cross-oriented combinations generate strong responses in situations where neither component in isolation is effective. Thus the cells respond to the configuration of the stimulus rather than its parts alone, so they can paradoxically respond well to an inner stimulus at 90° to their optimal in the presence of a cross-oriented outer stimulus. Thus, for many of these cells in V1, the strongest response is elicited by a particular configuration of stimuli (for example, Fig. 2). We suggest that these cells play a role in detecting the location and direction of a change in contour orientation associated, for example, with junctions or corners. □



Received 8 June; accepted 12 September 1995.

- Maffei, L. & Fiorentini, A. *Vision Res.* **16**, 1131–1139 (1976).
- Fries, W., Albus, K. & Creutzfeldt, O. D. *Vision Res.* **17**, 1001–1008 (1977).
- Nelson, J. I. & Frost, B. J. *Brain Res.* **139**, 359–365 (1978).
- Gilbert, C. D. & Wiesel, T. N. *Vision Res.* **30**, 1689–1701 (1990).
- DeAngelis, G. C., Robson, J. G., Ohzawa, I. & Freeman, R. D. *J. Neurophysiol.* **68**, 144–163 (1992).
- Grinvald, A., Lieke, E. E., Frostig, R. D. & Hildesheim, R. J. *Neurosci.* **14**, 2545–2568 (1994).
- Born, R. T. & Tootell, R. B. H. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7071–7075 (1991).
- Sillito, A. M. *J. Physiol.* **273**, 791–803 (1977).
- Orban, G. A., Kato, H. & Bishop, P. O. *J. Neurophysiol.* **42**, 818–832 (1979).
- Orban, G. A., Kato, H. & Bishop, P. O. *J. Neurophysiol.* **42**, 833–849 (1979).
- Ts'o, D. Y., Gilbert, C. D. & Wiesel, T. N. *J. Neurosci.* **6**, 1160–1170 (1986).
- Abeles, M. *J. Neurosci. Meth.* **5**, 317–325 (1982).
- Abeles, M. & Gerstein, G. L. *J. Neurophysiol.* **60**, 909–924 (1988).
- Palm, G., Aertsen, A. M. H. J. & Gerstein, G. L. *Biol. Cybern.* **59**, 1–11 (1988).
- Gray, C. M., König, P., Engel, A. K. & Singer, W. *Nature* **338**, 334–337 (1989).
- Engel, A. K., König, P., Gray, C. M. & Singer, W. *Eur. J. Neurosci.* **2**, 588–606 (1990).
- Gray, C. M., Engel, A. K., König, P. & Singer, W. *Eur. J. Neurosci.* **2**, 607–619 (1990).
- Bergin, J. R. & Julesz, B. *Nature* **303**, 696–698 (1983).
- Bergin, J. R. & Julesz, B. *IEEE Trans. Syst. M13*, 857–863 (1983).
- Keeble, D. R. T., Kingdom, F. A. A., Moulda, B. & Morgan, M. J. *Vision Res.* **35**, 1991–2006 (1995).
- Hubel, D. H. & Wiesel, T. N. *J. Neurophysiol.* **28**, 229–287 (1965).
- DeAngelis, G. C., Freeman, R. D. & Ohzawa, I. *J. Neurophysiol.* **71**, 347–374 (1994).
- Knierim, J. J. & Van Essen, D. C. *J. Neurophysiol.* **67**, 961–980 (1992).
- Sillito, A. M., Cudeiro, J. & Murphy, P. C. *Exp Brain Res.* **93**, 6–16 (1993).

ACKNOWLEDGEMENTS. We thank S. Tassen and S. Claxton for technical support. The support of the MRC is gratefully acknowledged.

## Topographical representations of mental images in primary visual cortex

Stephen M. Kosslyn<sup>\*†</sup>, William L. Thompson<sup>\*</sup>,  
Irene J. Kim<sup>\*</sup> & Nathaniel M. Alpert<sup>‡</sup>

<sup>\*</sup>Department of Psychology, Harvard University, Cambridge, Massachusetts 02138, USA

<sup>†</sup>Department of Neurology and <sup>‡</sup>Department of Radiology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

WE report here the use of positron emission tomography (PET) to reveal that the primary visual cortex is activated when subjects close their eyes and visualize objects. The size of the image is systematically related to the location of maximal activity, which is as expected because the earliest visual areas are spatially organized<sup>1–5</sup>. These results were only evident, however, when imagery conditions were compared to a non-imagery baseline in which the same auditory cues were presented (and hence the stimuli were controlled); when a resting baseline was used (and hence brain activation was uncontrolled), imagery activation was obscured because of activation in visual cortex during the baseline condition. These findings resolve a debate in the literature about whether imagery activates early visual cortex<sup>6–11</sup> and indicate that visual mental imagery involves 'depictive' representations, not solely language-like descriptions<sup>12–14</sup>. Moreover, the fact that stored visual information can affect processing in even the earliest visual areas suggests that knowledge can fundamentally bias what one sees.

We tested 12 right-handed male volunteers in five conditions while local cerebral blood flow was monitored using PET. The resting baseline task followed the procedure and instructions of ref. 10. Subjects were blindfolded with a dark cloth, and told to close their eyes, relax and to 'have it black in front of their mind's eye'. Such an uncontrolled rest state probably corresponds to many states that do not preclude some form of imagery. The other four conditions used the same type of auditory stimuli, which either were listened to passively or used as prompts to form images of previously memorized pictures (Fig. 1). Subjects received three imagery conditions, each of which consisted of four practice trials and 24 experimental trials. The imagery conditions differed only in the size at which images were formed, at

0.25, 4.0 or 16.0 degrees of visual angle. We compared activation in the three imagery conditions with each of the two baselines.

Here we focus entirely on activation in the medial occipital region. First, when activation in the resting baseline was subtracted from the imagery conditions the only occipital activation was in area 19 (only for medium-sized images), including the parieto-occipital sulcus and precuneus. This is very similar to the region reported previously with this baseline condition<sup>15</sup>; the *X Y Z* Talairach coordinates were  $-20 - 68 28$ . Second, Fig. 2 and Table 1 present the results of subtracting activation in the listening baseline from the three imagery conditions; we focus here on activation within 15 mm of the midline (the smoothness parameter used in analysis was 14 mm) in the occipital lobe, and consider only activations that had *Z* scores greater than 3.0. Locations are specified using the stereotactic coordinate system<sup>16</sup>. As is evident, we find very posterior activation only for the small images, and very anterior activation only for the

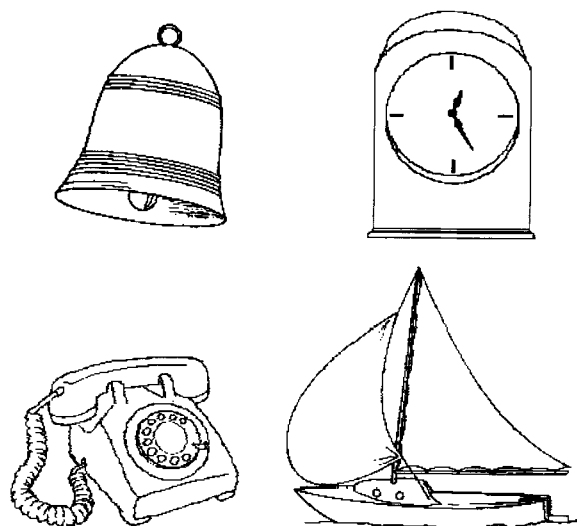
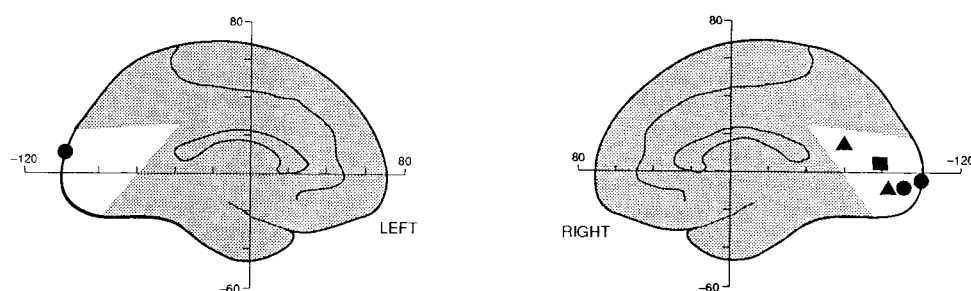


FIG. 1 Stimuli used in four of the five conditions. In the listening baseline condition, subjects received trials of the following sort. First they heard the name of a common object (such as 'anchor'), and 4 s later heard a spatial comparison term (such as 'right higher'), and then responded. One second later, another trial was presented. Subjects were told to close their eyes and respond as quickly as possible on hearing the comparison terms, alternating feet from trial to trial; they were told not to visualize anything during these 28 trials. The order of this baseline and the resting baseline was counterbalanced. Half of the subjects received the resting baseline at the beginning of the experiment, and half received it at the end of the experiment; in both cases, the listening baseline always preceded the first imagery condition (this was necessary because once subjects knew the meanings of the cues, they would be unlikely to listen passively). Immediately before the imagery conditions, subjects studied 28 pictures, adapted from ref. 19, as illustrated here. Each picture appeared on the screen when its name was read aloud by the computer. After memorizing the pictures, four spatial judgements were defined; for example, 'right higher' required them to decide whether the rightmost point was higher than the leftmost point. Pictures of objects (not used in the actual experiment) illustrated 'yes' and 'no' judgements. At the beginning of each set of imagery trials, subjects memorized the size of a square piece of cardboard taped to the screen. Subjects heard the same type of stimuli used in the listening baseline condition, but now closed their eyes and visualized each named picture at the size of the square. Subjects were urged to visualize each object at the correct size and to maintain the image at this size. When they heard the spatial comparison term, they evaluated the imaged picture as quickly and accurately as possible. Four different sets of 28 pictures were used, and counterbalancing ensured that the names and associated spatial comparison terms occurred equally often in the listening baseline and in each imagery condition in each order, and that the spatial comparison terms and yes/no responses appeared equally often in each condition.

FIG. 2 Left and right PET results in medial occipital cortex; points indicate the most activated pixel using statistical parametric mapping<sup>20</sup> with  $Z > 3.0$ . Small imagery (●), medium imagery (■), and large imagery (▲), all minus listening baseline. Activation for the smallest images was very posterior, -102 mm from the anterior commissure for left- and right-hemisphere area 18, which included area 17 in the right hemisphere, as well as -88 mm in a more medial portion of area 18. Activation for medium images was -79 mm in area 17 in the right hemisphere. Finally, activation for the largest images was -60 mm in a region that includes area 17 and the precuneus, and -83 mm in another part of area 17 and 18, which may have been activated by internal details of the images. The PET machine was a GE Scanditronix PC4096 15-slice whole-body tomograph in its stationary mode (see ref. 21). Contiguous slices were 6.5 mm apart (centre-to-centre; the axial field was 97.5 mm); the axial resolution was 6.0 mm full width at half maximum. Each subject was fitted with a custom moulded face mask (TRUE SCAN, Annapolis, MD), and his head was aligned relative to the CM (cantho-meatal) line. Nasal



cannulae were connected to a radiolabelled gas inflow, and a face mask, attached to a vacuum, was placed over the subject's nose. On each PET run, 20 measurements were taken: the first 3 were 10 s and the remaining 17 were 5 s long. The PET protocol was as follows: (1) the camera acquisition program was started (which measured residual background from past studies); (2) stimulus presentation began 15 s later, and the subject began the task; (3)  $^{15}\text{O}-\text{CO}_2$  gas was administered 15 s after this; and (4) 60 s later, scanning ended and the gas was stopped (for additional details, see experiment 2 of ref. 7).

largest images, with medium images activating intermediate regions. Inspection of the scans from each individual indicated that activation was in fact in the primary visual cortex.

We next compared the two baselines by subtracting blood flow during the listening baseline from that during the resting baseline. We found more blood flow during the resting baseline in area 17 (primary visual cortex),  $t(11) = 3.94$ ,  $P = 0.002$  (two-tailed) (Fig. 3). To examine the possible effects of practice in both baselines, we compared only the first condition for each subject, and found greater activation in area 17 during the resting baseline,  $F(1, 10) = 5.65$ ,  $P = 0.039$ ; we also compared only the second baseline condition for each subject, and again found more activation in area 17 during the resting baseline,  $F(1, 10) = 7.06$ ,  $P = 0.024$ . Moreover, when baseline and order were considered in a single analysis of variance, there was no interaction between the two factors,  $F < 1$ .

TABLE 1 Data showing blood flow under experimental conditions

| (a) Small-sized images – listening baseline  |     |      |    |         |
|----------------------------------------------|-----|------|----|---------|
|                                              | X   | Y    | Z  | Z-score |
| Left hemisphere regions                      |     |      |    |         |
| Area 18                                      | -13 | -102 | 8  | 3.20    |
| Right hemisphere regions                     |     |      |    |         |
| Area 18/17                                   | 15  | -102 | -4 | 3.40    |
| Area 18                                      | 6   | -88  | -8 | 3.42    |
| (b) Medium-sized images – listening baseline |     |      |    |         |
| Right hemisphere regions                     |     |      |    |         |
| Area 17                                      | 8   | -79  | 4  | 3.00    |
| (c) Large-sized images – listening baseline  |     |      |    |         |
| Midline                                      |     |      |    |         |
| Area 17*                                     | 2   | -83  | 0  | 3.25    |
| Area 18*                                     | -4  | -83  | -8 | 3.30    |
| Precuneus/17                                 | -2  | -60  | 16 | 3.09    |
| (d) Resting baseline – listening baseline    |     |      |    |         |
| Left hemisphere regions                      |     |      |    |         |
| Area 18/Cuneus                               | -7  | -102 | 8  | 3.79    |
| Area 17*                                     | -11 | -69  | 8  | 3.21    |
| Area 18*                                     | -6  | -67  | 4  | 3.39    |
| Right hemisphere regions                     |     |      |    |         |
| Area 19/18                                   | 12  | -92  | 24 | 3.65    |
| Area 18                                      | 6   | -81  | 0  | 3.60    |

Blood flow in the listening baseline condition was subtracted from blood flow in the small (a), medium (b), and large (c) visual mental imagery conditions and from blood flow in the resting baseline condition (d;  $Z > 3.0$ ).

\* Part of the same continuous area of activation.

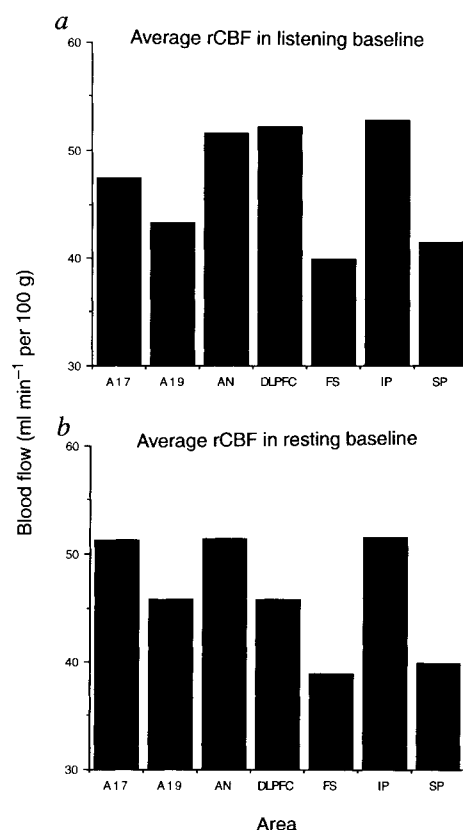


FIG. 3 Blood flow relative to the global mean for different brain areas during the two baseline conditions. Each brain was normalized to the same value,  $50 \text{ ml min}^{-1} \text{ per } 100 \text{ g}$ , and the relative flows in area 17 and other cortical areas were analysed relative to the mean. Activation in area 17 (A17) was average in the listening baseline (a), but above average in the resting baseline (b). The activation in area 17 was compared to the mean blood flow in the previously activated regions<sup>7</sup> of area (A19), the angular gyrus (AN), dorsolateral prefrontal cortex (DLPFC), the fusiform gyrus (FS), the inferior parietal lobe (IP), and the superior parietal lobe (SP), not the normalized value of  $50 \text{ ml min}^{-1} \text{ per } 100 \text{ g}$ ; thus, the mean blood flow in A17 (47.4) is not different ( $F < 1$ ) from the mean blood flow in those other brain areas (46.8) during the listening baseline, whereas the mean blood flow in A17 (51.3) is higher than that in the other brain areas (45.6) during the resting baseline ( $F = 59.8$ ;  $P = 0.0001$ ). Most of the same extrastriate regions were activated during each of the three sizes, but not all such areas were activated in common.

We also considered the variability associated with each baseline, and found comparable variances (5.91 and 7.65 for the listening and resting baselines, respectively). Moreover, we found more blood flow in area 17, relative to the global mean, than in the average of the other cortical areas we examined in the resting baseline condition (consistent with ref. 17), but not in the listening baseline condition, (see Fig. 3).

The present findings do not imply that visual images are stored in primary visual cortex. About 32 cortical areas have now been shown to process visual information in the macaque monkey brain, and the spatially organized early areas not only send information to areas later in the processing stream, but also receive information from those later areas<sup>18</sup>; indeed, the feedback connections are of comparable size to the feedforward connections<sup>4</sup>. Thus, it is possible that higher-level areas that actually store visual information can evoke activity in earlier areas.

The fact that the size of the image systematically affected the locus of activation rules out the possibility that hypometabolism in visual cortex artefactually produced the appearance of imagery activation when the listening baseline was used. Moreover, the use of three sizes eliminates the possibility that different sizes activated different brain areas (as could have occurred in ref. 7, in which two sizes were examined, and which also did not include any baseline conditions but rather examined activation only relative to the other size, and thereby did not allow precise localization of activity). Finally, the relatively high occipital activation during the resting baseline explains why previous researchers using this baseline failed to find evidence that imagery activates area 17. By using three sizes of pictures of common objects, comparing them to a baseline that had the same physical stimuli (but different, non-imagery, mental processing), and showing

that the three sizes induce activation along a continuum (but only when we compare them to the listening baseline), we circumvent the problems inherent in the earlier designs, thereby convincingly demonstrating that visual imagery activates primary visual cortex. □

Received 3 August; accepted 3 October 1995.

1. Daniel, P. M. & Whittridge, D. J. *Physiol.* **159**, 203–221 (1961).
2. Fox, P. T. et al. *Nature* **323**, 806–809 (1986).
3. Tootell, R. B. H., Silverman, M. S., Switkes, E. & De Vailois, R. L. *Science* **218**, 902–904 (1982).
4. Felleman, D. J. & Van Essen, D. C. *Cereb. Cortex* **1**, 1–47 (1991).
5. Fox, P. T. et al. *Nature* **323**, 806–809 (1986).
6. Damasio, H. et al. *Soc. Neurosci. Abstr.* **19**, 1603 (1993).
7. Kosslyn, S. M. et al. *J. cogn. Neurosci.* **5**, 263–287 (1993).
8. Menon, R. et al. in *Functional MRI of the Brain: A Workshop Presented by the Society of Magnetic Resonance in Medicine and the Society for Magnetic Resonance Imaging* 2nd edn (eds Le Bihan, D., Turner, R., Mosley, M. & Hyde, J.) (Society of Magnetic Resonance in Medicine Inc, Arlington, VA, 1993).
9. Charlot, V., Tzourio, M., Zilbovicius, M., Mazoyer, B. & Denis, M. *Neuropsychologia* **30**, 565–580 (1992).
10. Roland, P. E. & Guliyas, B. *Cereb. Cortex* **5**, 79–93 (1995).
11. Le Bihan, D. et al. *Proc. natn. Acad. Sci. U.S.A.* **5**, 11802–11805 (1993).
12. Pylshyn, Z. W. *Psychol. Bull.* **80**, 1–24 (1973).
13. Kosslyn, S. M. & Pomerantz, J. R. *Cogn. Psychol.* **9**, 52–76 (1977).
14. Kosslyn, S. M. *Image and Brain: The Resolution of the Imagery Debate* (MIT Press, Cambridge, MA, 1994).
15. Roland, P. E. & Friberg, L. J. *Neurophys.* **53**, 1219–1243 (1985).
16. Talairach, J. & Tournoux, P. *Co-planar Stereotaxic Atlas of the Human Brain* (trans. M. Rayport) (Thieme, New York, 1988).
17. Roland, P. E. *Brain Activation* (Wiley-Liss, New York, 1993).
18. Douglas, K. L. & Rockland, K. S. (1992) *Soc. Neurosci. Abstr.* **18**, 390 (1992).
19. Snodgrass, J. G. & Vanderwart, M. A. *J. exp. Psychol. hum. Learn. Mem.* **6**, 174–215 (1980).
20. Friston, K. J., Frith, C. D., Liddle, P. F. & Frackowiak, R. S. J. *J. Cereb. Blood Flow Metab.* **11**, 690–699 (1991).
21. Kops, E. R., Herzog, H. H., Schmid, A., Holte, S. & Feinendegen, L. E. *J. Comput. Assist. Tomogr.* **14**, 437–445 (1990).

ACKNOWLEDGEMENTS. We thank A. Loring, S. Weise and the MGH Cyclotron Unit for technical assistance.

## Recovery from spinal cord injury mediated by antibodies to neurite growth inhibitors

Barbara S. Bregman, Ellen Kunkel-Bagden,  
Lisa Schnell\*, Hai Ning Dai, Da Gao  
& Martin E. Schwab\*

Department of Cell Biology, Division of Neurobiology,  
Georgetown University Medical Center, 3900 Reservoir Road NW,  
Washington DC 20007, USA

\* Brain Research Institute, University of Zurich,  
August Forel-Strasse 1, CH-8029 Zurich, Switzerland

THERE is little axonal growth after central nervous system (CNS) injury in adult mammals. The administration of antibodies (IN-1) to neutralize the myelin-associated neurite growth inhibitory proteins leads to long-distance regrowth of a proportion of CNS axons after injury<sup>1–5</sup>. Our aim was: to determine if spinal cord lesion in adult rats, followed by treatment with antibodies to neurite growth inhibitors, can lead to regeneration and anatomical plasticity of other spinally projecting pathways; to determine if the anatomical projections persist at long survival intervals; and to determine whether this fibre growth is associated with recovery of function. We report here that brain stem–spinal as well as corticospinal axons undergo regeneration and anatomical plasticity after application of IN-1 antibodies. There is a recovery of specific reflex and locomotor functions after spinal cord injury in these adult rats. Removal of the sensorimotor cortex in IN-1-treated rats 2–3 months later abolished the recovered contact-placing responses, suggesting that the recovery was dependent upon the regrowth of these pathways.

Young adult (6–8 weeks of age, 150–200 g) Lewis rats ( $N=45$ ) received a mid-thoracic microsurgical spinal cord lesion (over hemisection, HX) interrupting the dorsal columns and corticospinal axons bilaterally and the lateral and ventral funiculi and intervening grey matter unilaterally as described previously<sup>6–8</sup>. At the time of spinal cord surgery, hybridoma cells secreting IN-1 antibody ( $N=23$ ) or secreting control (horseradish peroxidase, HRP) antibodies ( $N=22$ ) were placed into the parietal cortex<sup>1,2</sup>. Small slowly growing tumours formed under immune suppressive treatment (cyclosporin A) and secreted antibodies that reached the spinal cord through the cerebrospinal fluid<sup>1,2</sup>. Cyclosporin A was discontinued after two weeks; this led to the resorption of the hybridoma tumours within about a week. Additional rats received the spinal cord lesion without any further treatment ( $N=16$ , HX). Unlesioned age-matched rats ( $N=16$ ) served as controls. Spinal cord lesions and IN-1 treatment were performed in Zurich, rats were number coded and randomly mixed groups were shipped to Washington and analysed without knowledge of treatment history. Rats were allowed to recover for 4–6 weeks before behavioural training and qualitative and quantitative analysis of reflex and locomotor function<sup>6,7,9,10</sup>. At the completion of the behavioural analysis (3–4 weeks later), regeneration of corticospinal axons was assessed using anterograde tracing with horseradish peroxidase<sup>1–3,11</sup>. Serotonergic and noradrenergic pathways were visualized immunocytochemically<sup>10,12,13</sup>. The spinal cord lesion site was reconstructed serially for each of the animals. Animals in which the transverse extent of the lesion was smaller or larger than intended were excluded from the analysis. In order to determine the extent to which the regrowth of corticospinal axons contributed to the recovery of function in IN-1 treated animals, we removed the sensorimotor cortex bilaterally in a group of animals 3 months after the initial spinal cord injury (HX + IN-1,  $N=5$ ;

HX+HRP,  $N=4$ ) and in age-matched control rats without spinal cord lesions ( $N=4$ ).

The IN-1 treatment resulted in the growth of corticospinal axons around the lesion site and into the spinal cord caudal to the injury as described in earlier studies<sup>1,3</sup> and Fig. 1. The regenerated corticospinal axons persisted within the caudal spinal cord caudal to the lesion up to 12 weeks after the injury (the longest time point analysed). Corticospinal axons could be traced in both white matter and within the grey matter of the spinal cord caudal to the mid-thoracic lesion as far as the lumbar spinal cord segments in the IN-1 treated animals. Qualitatively, the projection is sparse; it appears that the proportion of corticospinal axons that grow long distances represents no more than 5–10% of the population which normally projects to the lumbar cord. The corticospinal projection in the rats without IN-1 treat-



FIG. 1 Camera lucida reconstruction of a spaced series of 50- $\mu$ m transverse sections of the spinal cord at and adjacent to the site of spinal cord lesion in an IN-1-treated animal. The most rostral section is to the upper left of the figure and the most caudal section to the lower right (total distance from first to last section drawn is 4.5 mm). Horseradish peroxidase was injected into the sensorimotor cortex bilaterally 48 h before perfusion; tissue was reacted for HRP-visualization using tetramethylbenzidine<sup>1,11</sup>. Dark areas and dots indicate the distribution of corticospinal fibres labelled with HRP. The corticospinal axons are in their normal position in the base of the dorsal columns in the most rostral sections, but the effects of the lesion are evident in the increased projection (both number and size) of axons as the lesion is reached. At its maximum transverse extent this lesion removed the dorsal column bilaterally, the grey matter and dorsolateral and lateral white matter on the right. Ventral white matter remains intact. IN-1 treatment permits the growth of corticospinal axons around the site of spinal cord injury. Note the appearance of corticospinal axons within an ectopic position in the white matter of the lateral funiculus in the sections at and caudal to the lesion site. Some axons resume a diffuse location within the dorsal columns caudal to the lesion. Corticospinal axons could be traced as far as the lumbar spinal cord segments in the IN-1-treated animals, and could be identified both in the white matter and within normal regions of termination in the grey matter of the lumbar spinal cord. The administration of IN-1 antibodies at the time of spinal cord injury leads to long-term growth of CS axons which is maintained for at least 3 months after the injury.

ment (HX, HX+HRP) was restricted to within 1–2 mm of the lesion site and did not reach lower thoracic or lumbar spinal cord levels at all; in fact, many of the cut corticospinal axons formed retraction bulbs rostral to the lesion, as demonstrated previously<sup>11</sup>. The administration of IN-1 antibodies at the time of the spinal cord injury leads to the growth of corticospinal axons below the level of injury and these axons are maintained up to 12 weeks later.

In order to determine whether IN-1 treatment is effective in increasing the growth of other descending projections, we have used immunocytochemical methods to visualize the distribution of the functionally very important raphespinal (serotonergic) and coeruleospinal (noradrenergic) axons after spinal cord injury and IN-1 treatment. As shown in Fig. 2, spinal cord hemisection in the adult dramatically decreased the projection of serotonergic axons to the spinal cord caudal and ipsilateral to the lesion<sup>12</sup> and control antibody did not alter this response (Fig. 2*b*). Treatment with IN-1 after spinal cord injury greatly increased the serotonergic (Fig. 2*c*) projection within the spinal cord caudal to the lesion in all treated animals ( $N=5$  in each group). Quantitative image analysis indicates a significant increase ( $t$ -test,  $P=0.001$ ) in the serotonergic projection within the lumbar spinal cord caudal and ipsilateral to the hemisection in the IN-1-treated animals (Fig. 2*d*). The serotonergic projection in the ventral horn of the lumbar spinal cord caudal and ipsilateral to the lesion in HX+IN-1-treated animals was more than 3 times that in the animals that had received the control antibody. A similar increase in axonal growth was also seen with noradrenergic projections (data not shown). Because intact serotonergic fibre tracts are present in the lateral and ventral funiculi of the unlesioned side of the spinal cord we cannot distinguish the relative contribution of regenerative growth of serotonergic fibres around the lesion site versus axonal sprouting of contralateral undamaged pathways to the remodelling of brain stem–spinal projections under the influence of IN-1.

Qualitative and quantitative analysis of reflex and locomotor function in these adult treated rats were performed. One aspect of sensorimotor reflex function that is known to be dependent upon the integrity of the corticospinal pathway is the contact-placing response<sup>14–17</sup>. In response to low threshold hair or light skin contact of the foot (without joint displacement) the animal lifts its limb and places it on a surface for support. In adult animals, spinal cord injury leads to a loss of contact-placing responses which does not recover<sup>9,14,15,17</sup>, whereas the re-routing of corticospinal axons which occurs after neonatal spinal cord injury leads to sparing of contact-placing responses<sup>9,14,15,17</sup>. Lesion only (HX) and control antibody-treated rats (HX+HRP) were identical; spinal cord lesion interrupting the corticospinal pathway permanently abolished contact placing (Fig. 3) although placing elicited by joint bending (proprioceptive placing) recovered (data not shown). In contrast, 80% of the IN-1-treated rats recovered contact-placing responses (Fig. 3). The recovered contact-placing response differed from the response observed in normal animals in that in IN-1-treated rats 40–60% of the trials elicited a placing response versus >90% of trials in normal animals. The kinematics of the movement revealed that in some of the IN-1-treated animals there was an exaggerated range of movement (hypermetria) in the recovered contact placing whereas in others the kinematics of the recovered movement was similar to that in normal animals. Following recovery (2–3 months) a group of rats which had completed the behavioural analysis underwent a subsequent bilateral cortical ablation to remove the source of the corticospinal fibres. In each of these five IN-1-treated animals, in which contact placing had recovered, the cortical ablation abolished contact placing permanently (6 weeks). Proprioceptive placing recovered in each of these rats by 1 week after the cortical ablation. The cortical ablation also abolished contact placing in the normal animals but had no effect on placing responses in the HX+HRP group (as in those animals contact placing never recovered after the

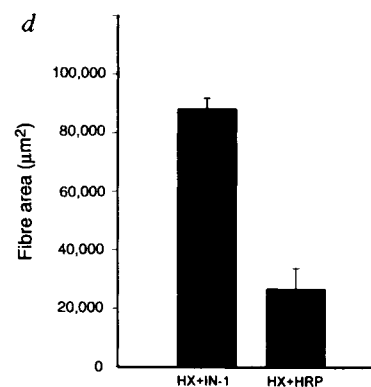
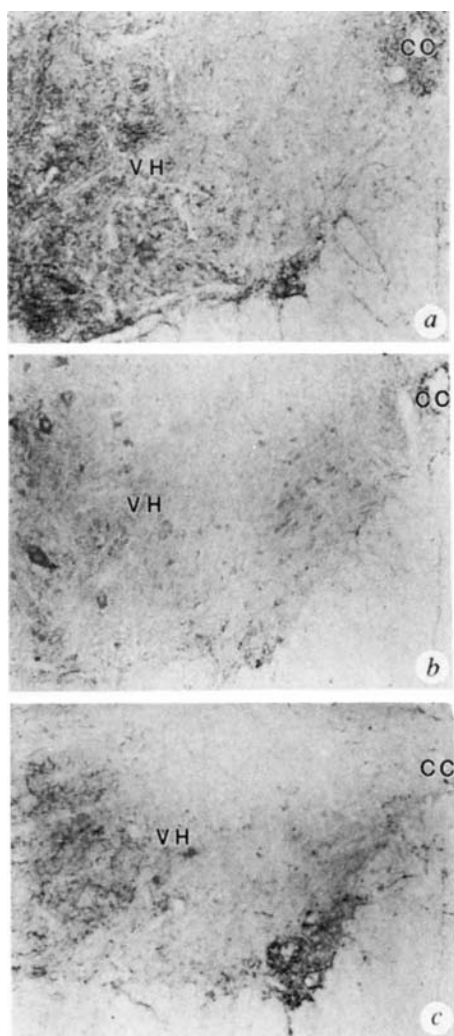
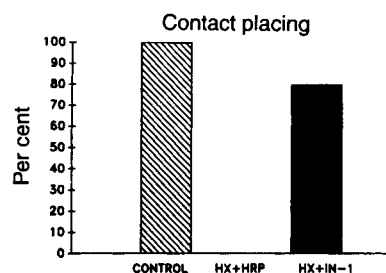


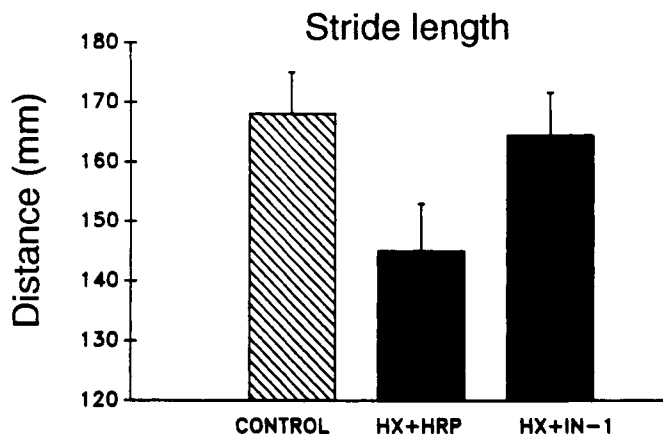
FIG. 2 IN-1 treatment increases the growth of brain stem–spinal axons after spinal cord injury. Representative transverse sections through the lumbar spinal cord ventral horn (35–40 mm caudal to the lesion, *b*, *c*) reacted immunocytochemically to visualize raphespinal serotonergic axons<sup>12,13</sup>. *a*, Normal (unlesioned) control; *b*, mid-thoracic 'over-hemisection', control (HRP) antibody treated; *c*, mid-thoracic 'over-hemisection', IN-1 antibody treated. VH, Ventral horn; CC, central canal. Sections from all 3 animals were processed simultaneously on the same slide for immunocytochemistry according to established procedures<sup>10,12,13</sup>; lesions in *b* and *c* were matched for transverse extent of damage and the lesion in both destroyed the dorsal columns bilaterally, dorsolaterally, laterally, and the ventral funiculus (and intervening grey matter) unilaterally; there was very slight damage to the contralateral ventral funiculus in both *b* and *c*. The increased density of the projection of the serotonergic axons within the ventral horn in the IN-1-treated animals (*c*) indicates that IN-1 treatment increases the anatomical plasticity of brain stem–spinal axons compared to lesion only or control antibody (*b*). Quantitative image analysis indicates a significant increase in the serotonergic axons within the spinal cord in the IN-1-treated animals (*d*). To estimate the density of the serotonergic projection we used a Zeiss/IBAS image analysis system to measure the area fraction occupied by 5-HT axons<sup>12</sup>. We systematically measured the serotonergic projection within a  $5.5 \times 10^6 \mu\text{m}^2$  region of L5 ventral horn grey matter caudal and ipsilateral to the lesion in HX + IN-1 ( $N=3$ ) and HX + HRP ( $N=3$ ) lesion-matched animals, processed simultaneously for immunocytochemistry, 4 sections per animal. The mean area fraction occupied by 5-HT fibres in IN-1-treated animals was  $88,170.4 \mu\text{m}^2$  and in HRP treated animals the mean was  $26,747.5 \mu\text{m}^2$ . The serotonergic innervation was significantly greater in the IN-1-treated group than in the control antibody-treated group ( $t$ -test;  $t = 7.867$ ,  $\text{d.f.} = 4$ ,  $P = 0.001$ ).

FIG. 3 In order to determine whether IN-1 treatment enhances recovery of function after spinal cord injury in adult animals we have used qualitative and quantitative analysis of reflex and locomotor function in adult rats with partial spinal cord lesions (over-hemisection, HX, as described above). The motor capacity of IN-1-treated rats (HX + IN-1) was compared with that of rats receiving lesions and control antibody (HX + HRP) or lesion only (HX, data not shown, but indistinguishable from the HX + HRP group) and with that of unlesioned age-matched control animals. All behavioural testing was conducted by individuals unaware of the experimental treatment status of the animals (randomly mixed groups of number-coded animals). At the completion of the behavioural studies, the transverse extent of the lesions were reconstructed serially from histological sections, the anatomical regrowth of descending pathways documented, the quantitative analysis of the behaviour was completed, and then the identity of the treatment groups revealed. Rats were tested for the presence of contact-placing responses following spinal cord lesion and IN-1 treatment. Contact-placing responses were tested with a force transducer to ensure that a low-threshold stimulus was delivered which involved only hair bend and skin contact (no joint bend). Lateral contact placing is abolished after spinal cord injury in lesion only and control antibody-treated animals, and does not recover. Lateral contact placing recovers in 80% of the animals treated with IN-1 ( $N=10$  in each group). HX- or HX + HRP-treated rats had no low-threshold placing responses. Placing elicited by higher-threshold proprioceptive (skin deformation, joint bending) was present in the HX + HRP-treated group. The recovery of contact placing responses after



IN-1 treatment may reflect the long-distance anatomical regrowth of corticospinal axons beyond the level of spinal cord injury. As described previously<sup>14–17</sup>, this reflex is dependent on a functional corticospinal connection. The observation that the frequency with which contact placing responses are elicited in the HX + IN-1 group compared to control animals and differences in the kinematics of the movement in some (not all) of the IN-1-treated animals may relate to the relatively small population of axons that project beyond the lesion site, variability in the nature of the ascending input which is re-established, or both. The loss of contact placing following cortical ablation suggests that the recovered movement was in fact dependent upon the regrowth of the corticospinal projections (either directly to the cord or indirectly through relays to brain stem nuclei, or both).





initial spinal cord lesion). These results suggest that the recovered movement in IN-1-treated animals was in fact dependent upon the regrowth of the corticospinal axons. Remodelling of cortical projections to brain stem nuclei (such as red nucleus, reticular nuclei) and plasticity of ascending sensory systems may also contribute to aspects of the motor recovery.

Control and treated animals were trained to cross runways for food and water rewards, and to walk on a treadmill<sup>6,7,9,10,18</sup>. One indication of the locomotor deficits resulting from spinal cord overhemisection was a significant permanent decrease in stride length after injury (Fig. 4). IN-1-treated animals recovered their stride length to levels equivalent to those observed in unlesioned animals (Fig. 4). In spite of the significant recovery of some aspects of locomotor function (stride length, for example) in the presence of IN-1, permanent deficits in other aspects of locomotor function were maintained. For example, the error rate of foot placement on grid runways for IN-1-treated and

FIG. 4 Locomotor behaviour was documented on videotapes and by footprint analysis<sup>6,7,9,10,18</sup>. Spinal cord injury in the adult leads to permanent deficits in locomotor function. After spinal cord injury, IN-1-treated animals recover locomotor function. Spinal cord injury in adult rats leads to a permanent decrease in stride length on the side of the injury but no change in stride length in the contralateral hindlimb. Application of tumours secreting control antibodies did not lead to recovery of stride length. In animals treated with IN-1, stride length in the hindlimb ipsilateral to the lesion recovers to levels identical to that in unlesioned control animals ( $N=9$  in each group). Analysis of variance indicates that there are significant differences ( $P<0.001$ ) in the treatment groups. Subsequent Bonferroni  $t$ -test indicates that the stride length in the HX + IN1 treatment group did not differ from that in control animals. Stride length in the HX + HRP group was significantly decreased ( $P<0.05$ ) from that in control and IN-1-treated groups. Statistical analysis indicates that the subsequent cortical ablation in a subset of animals that had recovered had no significant effect on stride length. This suggests that the pathways that influence stride length in these animals are not dependent upon cortical control. We suggest that improvement in locomotor function such as stride length after hemisection plus IN-1 treatment may be due to the reorganization of brain stem-spinal pathways. This anatomical plasticity may occur at spinal segmental levels, by long-distance regrowth of brain stem-spinal axons (analogous to that observed in the corticospinal pathways) or both. The current experiments demonstrate an increased projection of serotonergic axons within the spinal cord caudal and ipsilateral to the hemisection, but do not allow us to distinguish whether those fibres arise from increased axonal sprouting of contralateral pathways (which does not occur without IN-1 treatment) or by increased growth of the axons around the lesion site.

HRP-treated rats was identical (mean number of errors/crossing: IN-1 = 10.0, HRP = 10.3). This precise sensory-motor task may require a more complete regeneration of long-distance projections and greater remodelling of ascending sensory projections than could be achieved under the present conditions.

This study demonstrates that neutralization of myelin-associated neurite growth inhibitors in adult spinal cord lesioned rats can enhance regeneration and anatomical plasticity of corticospinal and brain stem-spinal pathways and mediate noticeable improvements in important aspects of locomotor function. They are in agreement with results in animals and humans with sub-total spinal cord lesions that show substantial recovery of specific functions mediated presumably by relatively small numbers of descending fibres influencing local spinal circuits<sup>19-25</sup>. Taken together, these observations suggest that in the future, specific intervention strategies could improve the functional outcome of spinal cord injury in humans by increasing anatomical plasticity and regeneration. □

Received 21 April; accepted 3 October 1995.

1. Schnell, L. & Schwab, M. E. *Nature* **343**, 269-272 (1990).
2. Schnell, L. & Schwab, M. E. *Eur. J. Neurosci.* **5**, 1156-1171 (1993).
3. Schnell, L., Schneider, R., Kolbeck, R., Barde, Y.-A. & Schwab, M. E. *Nature* **367**, 170-173 (1994).
4. Cadelli, D. & Schwab, M. E. *Eur. J. Neurosci.* **3**, 825-832 (1991).
5. Weibel, D., Cadelli, D. & Schwab, M. E. *Brain Res.* **642**, 259-266 (1994).
6. Kunkel-Bagden, E. & Bregman, B. S. *Expl Brain Res.* **81**, 25-34 (1990).
7. Kunkel-Bagden, E., Dai, H. N. & Bregman, B. S. *Expl Neurol.* **119**, 153-164 (1993).
8. Bregman, B. S. & Reier, P. J. *J. comp. Neurol.* **244**, 86-95 (1986).
9. Kunkel-Bagden, E., Dai, H. N. & Bregman, B. S. *Expl Neurol.* **116**, 40-51 (1992).
10. Bregman, B. S. *et al. Expl Neurol.* **123**, 3-16 (1993).
11. Bregman, B. S., Kunkel-Bagden, E., McAtee, M. & O'Neill, A. J. *comp. Neurol.* **282**, 355-370 (1989).
12. Bregman, B. S. *Dev Brain Res.* **34**, 245-263 (1987).
13. Bregman, B. S. *Dev Brain Res.* **34**, 265-279 (1987).
14. Bregman, B. S. & Goldberger, M. E. *Dev Brain Res.* **9**, 119-135 (1983).
15. Bregman, B. S. & Goldberger, M. E. *Science* **217**, 553-555 (1982).

16. Amassian, V. E. & Ross, R. J. *J. Physiol.* **74**, 165-184 (1978).
17. Bregman, B. S. & Goldberger, M. E. *Dev Brain Res.* **9**, 137-154 (1983).
18. Goldberger, M. E., Bregman, B. S., Vierk, C. J. & Brown, M. *Expl Neurol.* **107**, 113-117 (1990).
19. Eidelberg, E., Straehley, D., Erspamer, R. & Watkins, C. J. *Expl Neurol.* **56**, 312-322 (1977).
20. Alstermark, B., Lundberg, A., Pettersson, L.-G., Tantisira, B. & Walkowska, M. *Neurosci. Res.* **5**, 66-73 (1987).
21. Barbeau, H. & Rossignol, S. *Brain Res.* **546**, 250-260 (1991).
22. Vilensky, J. A., Moore, A. M., Eidelberg, E. & Waiden, J. G. *J. Motor Behav.* **24**, 288-296 (1992).
23. Wernig, A. & Muller, S. *Paraplegia* **30**, 229-236 (1992).
24. Dietz, V., Colombo, G. & Jensen, L. *Lancet* **344**, 1260-1263 (1994).
25. Wernig, A., Muller, S., Nanassy, A. & Cagol, E. *Eur. J. Neurosci.* **7**, 823-829 (1995).

ACKNOWLEDGEMENTS. We are extremely grateful to Ellen Kunkel-Bagden for her contributions to this work from its inception and we dedicate this manuscript to her memory. This work was supported by the Sheila Van Damm collaborative research grant from the International Spinal Research Trust. Support was also received from the International Research Institute for Paraplegia/Paul Schiller Foundation, Regeneration Pharmaceuticals, Tarrytown, NY, and the National Institute of Neurological Disorders and Stroke of the National Institutes of Health.

# A switch between two modes of synaptic transmission mediated by presynaptic inhibition

Melissa J. Coleman<sup>\*†‡</sup>, Pierre Meyrand<sup>§</sup>  
& Michael P. Nusbaum<sup>\*†||</sup>

<sup>\*</sup> University of Alabama at Birmingham, Department of Physiology and Biophysics and Neurobiology Research Center, 1719 6th Avenue South, Birmingham, Alabama 35294, USA

<sup>†</sup> University of Pennsylvania School of Medicine, Department of Neuroscience, 215 Stemmler Hall, Philadelphia, Pennsylvania 19104, USA

<sup>§</sup> Laboratoire de Neurobiologie et Physiologie Comparées, Université de Bordeaux I et CNRS, Place du Dr. Peyneau, 33120 Arcachon, France

<sup>‡</sup> Present address: Department of Biology, Brandeis University, Waltham, Massachusetts 02254, USA

<sup>||</sup> To whom correspondence should be addressed

**PRESYNAPTIC inhibition reduces chemical synaptic transmission in the central nervous system between pairs of neurons<sup>1-4</sup>, but its role(s) in shaping the multisynaptic interactions underlying neural network activity are not well studied. We therefore used the crustacean stomatogastric nervous system to study how presynaptic inhibition of the identified projection neuron, modulatory commissural neuron 1 (MCN1), influences the MCN1 synaptic effects on the gastric mill neural network. Tonic MCN1 discharge excites gastric mill network neurons and activates the gastric mill rhythm<sup>5,6</sup>. One network neuron, the lateral gastric (LG) neuron, presynaptically inhibits MCN1 and is electrically coupled to its terminals<sup>5,6</sup>. We show here that this presynaptic inhibition selectively reduces or eliminates transmitter-mediated excitation from MCN1 without reducing its electrically mediated excitatory effects, thereby switching the network neurons excited by MCN1. By switching the type of synaptic output from MCN1 and, hence, the activated network neurons, this presynaptic inhibition is pivotal to motor pattern generation.**

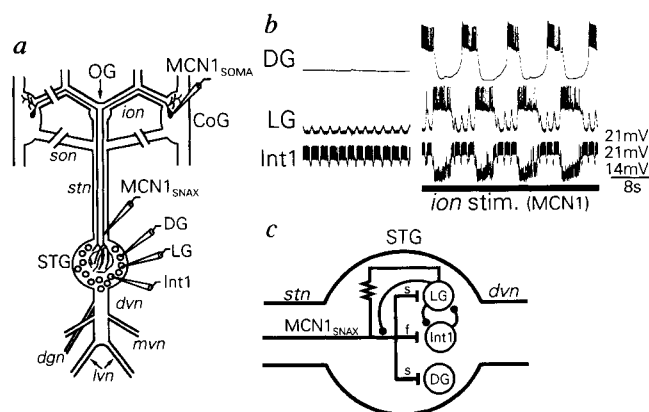
**FIG. 1.** *a*, Schematic of the isolated crab stomatogastric nervous system (STNS), including the intracellular recording sites of MCN1, LG, Int1 and DG. Interruptions in the *sons* and *ions* indicate nerve transections. Note that there is a single MCN1 in each CoG<sup>6,12</sup>. Each MCN1 influences the gastric mill neurons independently, and they do not influence each other<sup>6</sup>. Abbreviations: CoG, commissural ganglion; DG, dorsal gastric neuron; *dgn*, dorsal gastric nerve; *dvn*, dorsal ventricular nerve; Int1, interneuron 1; *ion*, inferior oesophageal nerve; LG, lateral gastric neuron; *lvn*, lateral ventricular nerve; MCN1<sub>SNAX</sub>, stomatogastric nerve axon of modulatory commissural neuron 1; MCN1<sub>SOMA</sub>, soma of MCN1; *mvn*, medial ventricular nerve; OG, oesophageal ganglion; *son*, superior oesophageal nerve; STG, stomatogastric ganglion; *stn*, stomatogastric nerve. *b*, Extracellular *ion* stimulation selectively activates MCN1, thereby eliciting a gastric mill rhythm in which LG bursts in alternation with Int1 and DG. Left, Before MCN1 stimulation, LG and DG were not active, whereas Int1 activity was time-locked to the pyloric rhythm<sup>10</sup>. The Int1 activity produced the rhythmic inhibition in LG. Right, Activation of MCN1 (15 Hz; black bar) elicited a gastric mill rhythm. Most hyperpolarized membrane potential: LG, -58 mV; Int1, -46 mV; DG, -64 mV. *c*, Schematic diagram of synaptic interactions within the STG. MCN1 produces fast (f: T-bar) chemical excitation of Int1 and slow (s: T-bars) chemical excitation of LG and DG. LG inhibits the MCN1 terminals (filled circle), and is also electrically coupled to them (resistor symbol)<sup>5,6</sup>. LG and Int1 are reciprocally inhibitory<sup>7,13</sup>.

**METHODS.** Adult male *Cancer borealis* ( $N=54$ ) were obtained from commercial suppliers. *C. borealis* saline: (in mM) NaCl, 440; KCl, 11; MgCl<sub>2</sub>, 26; CaCl<sub>2</sub>, 13; Trizma base, 10; and maleic acid, 5 (pH 7.4–7.6). When Mn<sup>2+</sup> (15 mM) was added to the saline, we reduced Ca<sup>2+</sup> levels (1.3 mM). Maintenance and dissection of crabs, as well as electrophysiological recording methods, were as described previously<sup>2-7</sup>.

The gastric mill rhythm is generated by a neural network located in the stomatogastric ganglion (STG)<sup>7-10</sup>. In the crab, *Cancer borealis*, this rhythm is influenced by projection neurons, including MCN1, that innervate the STG from the commissural ganglia (CoGs; Fig. 1*a*)<sup>6,11-13</sup>. MCN1 excites the gastric mill neurons LG, interneuron 1 (Int1) and the dorsal gastric (DG) neuron, but tonic MCN1 stimulation elicits a gastric mill rhythm in which LG impulse bursts alternate with those in Int1 and DG<sup>6</sup> (Fig. 1*b*). LG and Int1 burst in alternation, in part because they synaptically inhibit each other<sup>7,13</sup> (Fig. 1*b, c*). The alternation between LG and DG results from the LG-mediated presynaptic inhibition of the MCN1 terminals within the STG<sup>6</sup>. We documented the presence of this presynaptic inhibition by recording from the stomatogastric nerve axon (SNAX) of MCN1 (MCN1<sub>SNAX</sub>; Fig. 1*a*)<sup>5,6</sup>. However, MCN1<sub>soma</sub>-initiated action potentials propagate past the MCN1<sub>SNAX</sub> recording site before they are affected by the presynaptic inhibition. Therefore, we use the DG neuron as a reporter of MCN1 chemical synaptic transmission in the STG<sup>6</sup>. DG is an effective reporter because it is chemically excited by MCN1 and it is not synaptically influenced by any STG neurons.

A central aspect of the mechanism whereby MCN1 activity elicited the gastric mill rhythm that remained to be determined involved how each LG burst was generated, given that MCN1 was the only source of excitation to LG in these experiments and LG activity inhibited the STG terminals of MCN1. We found recently that LG in fact continues to receive excitatory postsynaptic potentials (e.p.s.ps) from MCN1 while LG is inhibiting this neuron ( $N=33$  preparations; Fig. 2*a*). This suggested that these e.p.s.ps resulted from the electrical connection between LG and MCN1. This possibility was supported by the fact that the MCN1-elicited e.p.s.p. was initiated earlier in LG than in Int1, by  $6 \pm 0.5$  ms ( $N=9$  preparations). Moreover, when all chemical transmission was suppressed, the e.p.s.p. was eliminated in Int1 but persisted with unchanged amplitude in LG ( $N=6$  preparations; Fig. 2*b*).

We also selectively eliminated chemical transmission from the STG terminals of MCN1 by impaling MCN1<sub>SNAX</sub> with a micro-electrode filled with potassium acetate (KAcetate) instead of potassium chloride (KCl). KAcetate eliminates chemical transmission from individual SNAXes while leaving electrical



Intracellular current injection was by single electrode, discontinuous current clamp (3 kHz sample rate; Axoclamp 2, Axon Instruments). Statistical analysis was performed using SigmaStat, version 1.0 (Jandel Scientific). Data were collected from isolated STNS preparations in which both *sons* and *ions* were transected, except for Fig. 2*a* where one *ion* was left intact. In experiments where extracellular *ion* stimulation was used to elicit MCN1 action potentials, MCN1 was the only activated neuron that influenced the STG because hyperpolarizing MCN1<sub>SNAX</sub> removed all the effects of *ion* stimulation ( $N=9$  preparations)<sup>20</sup>.

coupling intact. Under this condition, the e.p.s.p. again persisted in LG, but not in Int1 ( $N=5$  preparations; Fig. 2c; see also Fig. 3). We therefore conclude that the MCN1-mediated e.p.s.p. in LG results from electrical coupling. The persistence of electrical e.p.s.ps indicates that presynaptic inhibition does not eliminate the MCN1 action potentials in the STG, although the effective mechanism underlying this inhibition remains to be elucidated<sup>2,14-18</sup>.

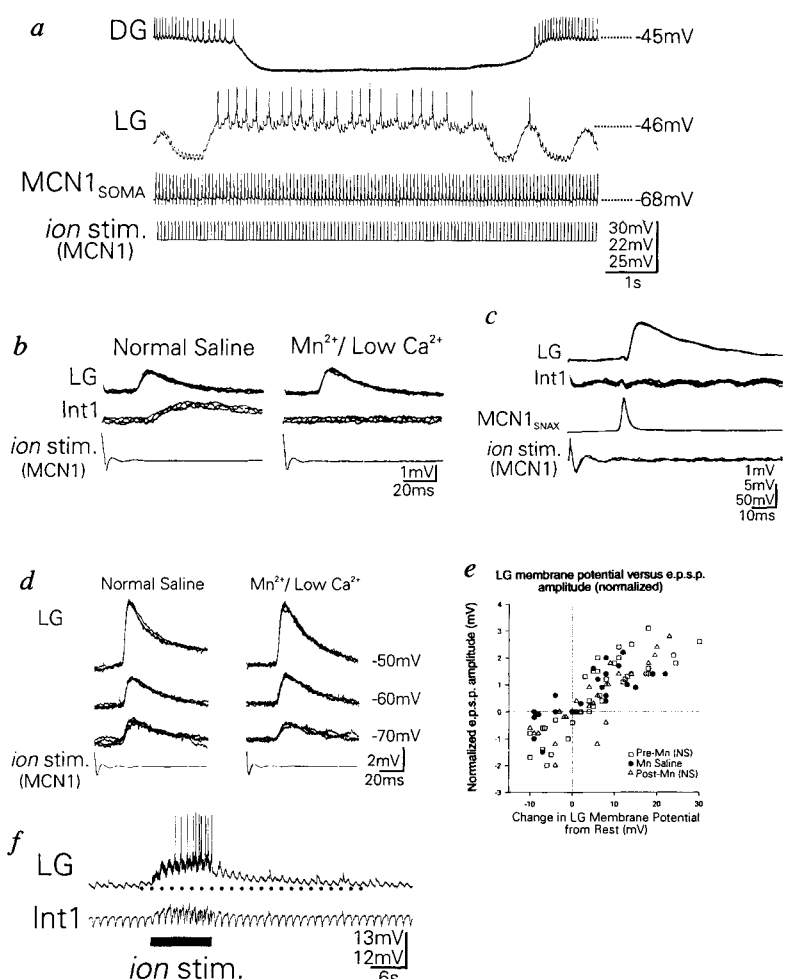
Interestingly, the amplitude of these electrical e.p.s.ps increased with depolarization and decreased with hyperpolarization of LG (Fig. 2d). Consequently, their amplitude was roughly double at membrane potentials near the peak versus the trough of the gastric mill-timed LG oscillations (Fig. 2e). This voltage dependence persisted when all chemical transmission was suppressed ( $N=11$  preparations; Fig. 2d, e). Voltage-dependent electrical coupling has been documented previously<sup>19</sup>, but its functional role is not well understood. Here, one likely function is to increase electrical excitation to LG while LG is presynaptically inhibiting its chemical excitatory input from MCN1. Additionally, the reduced amplitude of these e.p.s.ps at hyperpolarized levels will diminish their ability to oppose the synaptic inhibition from Int1, which is responsible for the LG interburst interval.

In addition to eliciting electrical e.p.s.ps in LG (Fig. 2b-d), MCN1 elicits a transmitter-mediated excitation of LG (Fig. 2f).

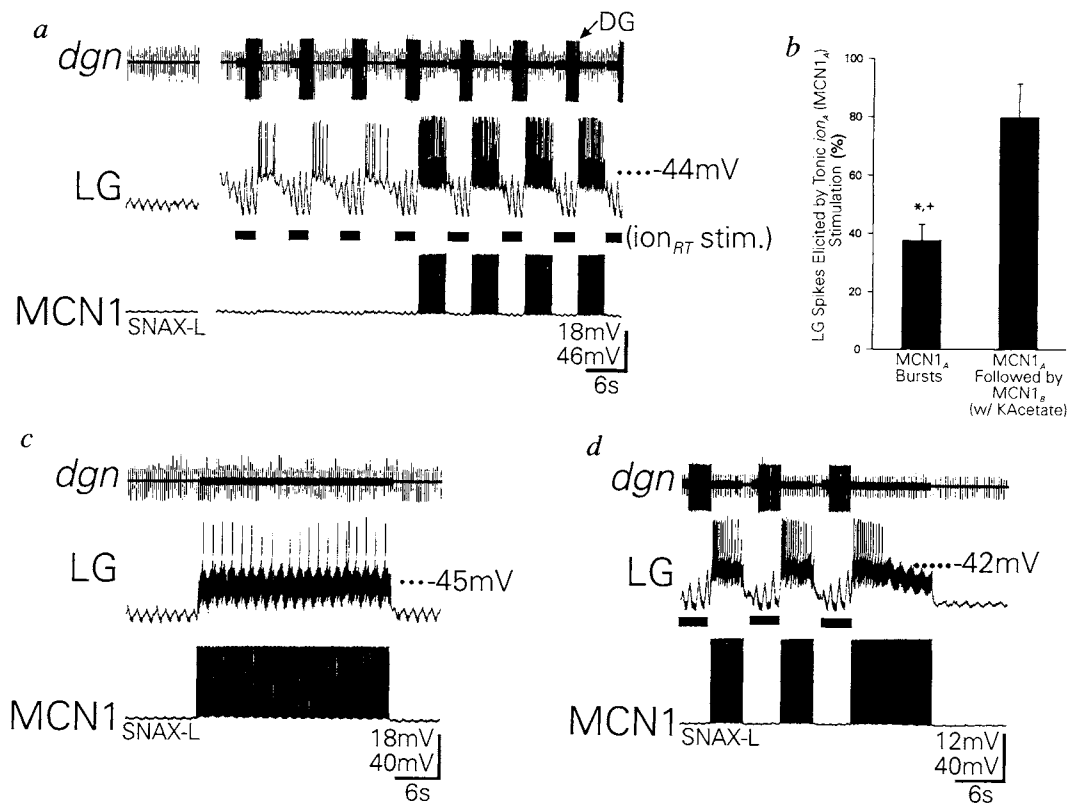
To show this excitation in Fig. 2f without interference from Int1 inhibition of LG, we hyperpolarized Int1 to prevent its activation by MCN1 stimulation. This transmitter-mediated excitation of LG develops slowly and, during the gastric mill rhythm, it enables LG to escape from Int1 inhibition<sup>20</sup>. Following termination of MCN1 activity, the underlying depolarization of LG decays slowly (Fig. 2f). In contrast, the purely electrical excitation of LG does not outlast MCN1 activity (see Fig. 3c).

We next examined whether the MCN1-mediated electrical e.p.s.ps and long-lasting excitation both contributed to each gastric mill rhythm-timed LG burst. Therefore, we mimicked the alternating chemical and electrical transmission from a tonically active MCN1 that is a consequence of the rhythmic, LG-mediated presynaptic inhibition. To accomplish this, we alternated rhythmic stimulation of an unaltered MCN1<sub>RIGHT</sub> with stimulation of MCN1<sub>LEFT</sub>, whose chemical transmission was selectively suppressed with KAcetate. Rhythmic stimulation of MCN1<sub>RIGHT</sub> alone elicited bursts in Int1 and DG that were comparable to their gastric mill rhythm-timed activity (Fig. 3a). Each Int1 burst produced the fast, rhythmic inhibition in LG. Int1 consistently is activated sooner than LG in response to MCN1 stimulation ( $N=13/13$  preparations). Following each MCN1<sub>RIGHT</sub> stimulation, LG fired a few action potentials (Fig. 3a). This LG activity is considerably weaker than its activity during a gastric mill rhythm, but stronger than that resulting

FIG. 2 MCN1 elicits electrotonic e.p.s.ps and slow, transmitter-mediated excitation of LG. **a**, Activation of MCN1, via *ion* stimulation (15 Hz), elicited the gastric mill rhythm, including alternating bursting in LG and DG. Each *ion* stimulus elicited an action potential in MCN1, which is recorded in MCN1<sub>SOMA</sub>. Each MCN1 action potential produced an e.p.s.p. in LG. These e.p.s.ps persisted during the time that LG inhibited MCN1, as is evident by the termination of the DG burst. **b**, Superimposed oscilloscope sweeps, triggered by *ion* stimulus (1 Hz), show that the MCN1-elicited e.p.s.p. in LG persists when chemical transmission is suppressed. Left, In normal saline, each MCN1 action potential elicits a fixed latency e.p.s.p. in LG and Int1. Right, In the same preparation, in  $Mn^{2+}$ /low  $Ca^{2+}$  saline, only the e.p.s.p. in LG persists. The e.p.s.p. in Int1 returned during the saline wash (not shown). Membrane potentials (left and right): Int1, -58 mV; LG, -61 mV. **c**, Superimposed sweeps showing that the MCN1-elicited e.p.s.p. in LG persists when MCN1-mediated chemical transmission is selectively eliminated by recording from MCN1<sub>SNAX</sub> with a KAcetate-filled electrode. Each *ion* stimulation (1 Hz) produced an MCN1 action potential, recorded in MCN1<sub>SNAX</sub>, that elicited a fixed latency e.p.s.p. in LG but no e.p.s.p. in Int1. In the same experiment, stimulation of the contralateral MCN1 produced a fixed latency e.p.s.p. in Int1 and LG (not shown). Saline contained high concentrations of  $Ca^{2+}$  (26 mM) to increase the amplitude of any chemically mediated e.p.s.ps. In this saline, larger e.p.s.ps occurred in Int1, but not in LG. Acetate ions are presumably interfering with transmitter release by altering intracellular pH near the MCN1 terminals, and thereby reducing  $Ca^{2+}$  influx<sup>30</sup>. Membrane potentials: MCN1<sub>SNAX</sub>, -58 mV; Int1, -55 mV; LG, -52 mV. **d**, The amplitude of the MCN1-mediated e.p.s.p. onto LG is voltage-dependent. Left, Composite figure showing superimposed sweeps in which each MCN1 action potential produced a fixed latency e.p.s.p. in the LG neuron. Depolarization of the LG membrane potential increased the e.p.s.p. amplitude. Right, The voltage-sensitivity persists in  $Mn^{2+}$ /reduced  $Ca^{2+}$  saline. This voltage sensitivity may result from gap junctional characteristics, because neither the LG nor MCN1<sub>SNAX</sub> input resistance changes across the studied voltage range (-40 mV to -90 mV)<sup>20</sup>. All recordings are from the same LG neuron impalement. **e**, Cumulative results showing the voltage-dependence of the electrical e.p.s.p. in LG ( $N=4$  preparations;  $n=5$  MCN1s). Each data point represents the normalized e.p.s.p. amplitude (e.p.s.p. at  $V_{holding}$ -e.p.s.p. at  $V_{rest}$ ) as a function of the change in LG membrane potential ( $V_{holding}-V_{rest}$ ), normalized to the resting potential. In these 5 experiments, the LG resting potential ranged from -58 mV to -66 mV. **a**, **b**, **c** and **d** are all from separate experiments. **f**, MCN1 stimulation (15 Hz) produces a slow, transmitter-



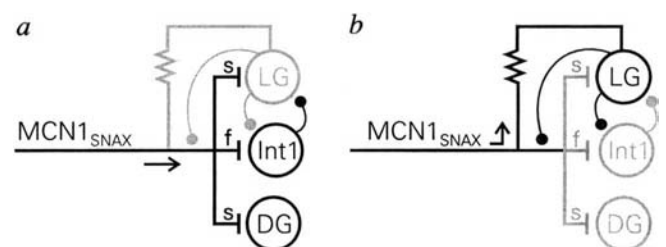
mediated excitation of LG in addition to the electrical e.p.s.ps. The depolarization underlying this excitation decays slowly after MCN1 activity is terminated. Int1 was hyperpolarized with constant amplitude current injection to eliminate its inhibitory influence on LG. Most hyperpolarized membrane potentials: LG, -58 mV; Int1, -76 mV.



**FIG. 3** MCN1-mediated electrical excitation contributes significantly to generation of each LG burst. *a*, Stimulation of  $ion_{RIGHT}$  (chemical transmission intact) in gastric mill rhythm-like bursts (bars; burst duration: 3.5 s; intraburst frequency: 15 Hz) enables MCN1<sub>RIGHT</sub> to excite Int1 (reflected in the increased amplitude, rhythmic inhibitory input to LG) and DG (*dgn*) rhythmically. Each stimulus duration corresponded to the duration of the Int1 burst/LG interburst interval during the gastric mill rhythm elicited by tonic  $ion_{RIGHT}$  stimulation. After each  $ion_{RIGHT}$  stimulation, each stimulation was followed by  $ion_{LEFT}$  stimulation (15 Hz for 3 s), which activated the KAcetate-filled MCN1<sub>SNAX-LEFT</sub>. There was a 500 ms delay between each  $ion_{RIGHT}$  and  $ion_{LEFT}$  stimulation. Membrane potential: MCN1<sub>SNAX-LEFT</sub>, -54 mV. *b*, Comparison of the number of LG spikes per burst during a gastric mill rhythm elicited by tonic stimulation of MCN1<sub>A</sub> (in which chemical transmission was intact) with (1) rhythmic stimulation of MCN1<sub>A</sub>, and (2) alternating stimulations of MCN1<sub>A</sub> and the contralateral, KAcetate-filled MCN1<sub>B</sub>. The number of LG spikes elicited by rhythmic MCN1<sub>A</sub> stimulation, and by alternating stimulation of MCN1<sub>A</sub> and MCN1<sub>B</sub> (chemical transmission suppressed) is expressed

as the percentage of the number of LG spikes elicited by tonic stimulation of MCN1<sub>A</sub> (mean  $\pm$  s.e.m.;  $N=4$  preparations). \* $P<0.05$  versus tonic MCN1<sub>A</sub> stimulation; \* $P<0.05$  versus MCN1<sub>A</sub> followed by MCN1<sub>B</sub> stimulation (Kruskal-Wallis one-way analysis of variance on ranks, followed by a pairwise multiple comparison procedure: Student-Newman-Keuls method). *c*, KAcetate suppresses chemical transmission from MCN1<sub>SNAX</sub>. MCN1<sub>SNAX-LEFT</sub> was recorded intra-axonally with a KAcetate-filled microelectrode. Stimulation of  $ion_{LEFT}$  (15 Hz) activated MCN1, eliciting e.p.s.ps and a weak activation of LG, but it did not activate DG. There was also no activation of Int1, shown by the absence of inhibitory synaptic input to LG. Small amplitude 'unit' in *dgn* is artefact from  $ion$  stimulation. *a* and *c* are from the same preparation. *d*, Electrical transmission from MCN1 to LG supports a maximal LG burst duration comparable to that occurring during the gastric mill rhythm. Experimental procedure is the same as *a*, except that following the final rhythmic activation of the unaltered MCN1<sub>RIGHT</sub> ( $ion_{RIGHT}$  stimulation, black bars), activation of the KAcetate-filled MCN1<sub>SNAX-LEFT</sub> was extended to 10 s. Intraburst firing frequency: MCN1<sub>RIGHT</sub>, 15 Hz; MCN1<sub>LEFT</sub>, 20 Hz. Membrane potential: MCN1<sub>SNAX-LEFT</sub>, -63 mV.

**FIG. 4** Presynaptic inhibition switches both the effective mode of MCN1 transmission and the targets activated by MCN1. *a*, When LG is not active (indicated by grey shading), MCN1 provides a fast chemical excitation to Int1 and a slow chemical excitation to DG and LG. The fast excitation of Int1 enables it to be activated in advance of LG. This further delays the LG burst onset because Int1 inhibits LG. When LG is relatively hyperpolarized, the electrical e.p.s.ps that it receives from MCN1 are small. The inhibition of LG allows MCN1 to continue releasing transmitter and, as a result, DG is also activated. Eventually, the slow chemical excitation that LG receives from MCN1 enables it to escape from the Int1-mediated inhibition and LG begins to fire action potentials. *b*, When LG is activated, it inhibits Int1 and presynaptically inhibits MCN1<sub>SNAX</sub>. The presynaptic inhibition reduces or eliminates the MCN1-mediated excitation of Int1 and DG, and these neurons stop firing. The MCN1-mediated chemical excitation of LG is also terminated, but the LG membrane potential repolarizes slowly, remaining depolarized for several seconds. Its maintained depolarization during the first 3–5 s after LG begins firing increases the strength of the voltage-dependent electrical e.p.s.ps that it continues to receive from MCN1. This enables



LG to continue firing action potentials. Eventually, the LG membrane potential repolarizes, reducing the effectiveness of the electrical e.p.s.ps and terminating the LG burst. This removes the presynaptic inhibition of MCN1, enabling MCN1 to again release transmitter and activate Int1 and DG. In *a* and *b*, black indicates active neurons, and grey indicates inactive neurons. Symbols are the same as in Fig. 1c.

from isolated activation of MCN1<sub>LEFT</sub> (Fig. 3b, c). To determine whether the MCN1 electrical excitation of LG was sufficient to rescue the LG bursts, we then stimulated the KAacetate-filled MCN1<sub>LEFT</sub> after each MCN1<sub>RIGHT</sub> stimulation ( $N=7$  preparations; Fig. 3a). This enhanced each LG burst sufficiently to make it equivalent to those occurring during the gastric mill rhythm (Fig. 3b).

The long-lasting depolarization of LG contributes to LG burst generation by increasing the effectiveness of the electrical e.p.s.ps. The time course of this depolarization also influences the duration of each LG burst. As is evident in Fig. 3d, when the duration of the MCN1<sub>LEFT</sub> activity is extended so that it outlasts the normal LG burst duration, the electrical e.p.s.ps only activate LG while it remains sufficiently depolarized by the long-lasting excitation ( $N=3$ ). Also consistent with a diminishing excitation of LG is the fact that the mean intraburst firing frequency of LG during the gastric mill rhythm is significantly faster during the first second of each burst than during the final second ( $8.4 \pm 1.1$  Hz versus  $4.9 \pm 1.5$  Hz;  $N=6$ ;  $P \leq 0.01$ , Wilcoxon Signed Rank Test).

These results indicate that the electrotonic excitation from MCN1 to LG contributes significantly to LG neuron activity. Thus, by selectively inhibiting chemical transmission, the presynaptic inhibition from LG switches the effective mode of MCN1 synaptic transmission. This enables MCN1 to excite, alternately, different subsets of gastric mill neurons and thereby elicit the gastric mill rhythm (Fig. 4).

In many systems, tonic activation of projection neurons, such as MCN1, or application of their modulatory transmitters is sufficient to evoke rhythmic neural network activity<sup>11,21,24</sup>. This has suggested that modulatory inputs often do not provide timing cues to the activated network. However, as shown here, in some systems there are local presynaptic events that alter the output from the distant terminals of such neurons, enabling them to have rhythmic influences on their network targets that are not evident from intra-somatic recordings. Recent studies have highlighted the presence of local properties in dendrites<sup>25,26</sup> as well as presynaptic influences in other rhythmic neural networks<sup>27-29</sup>. Therefore, similar presynaptic events are likely to shape neuronal interactions in all nervous systems. □

Received 18 August; accepted 10 October 1995.

1. Clarac, F., El Manira, A. & Cattaert, D. *Curr. Opin. Neurobiol.* **2**, 764-769 (1992).
2. Watson, A. H. D. *Comp. Biochem. Physiol. A* **103**, 227-239 (1992).
3. Rudomin, P., Quevedo, J. & Eguibar, J. R. *Curr. Opin. Neurobiol.* **3**, 997-1004 (1993).
4. Nusbaum, M. P. *Curr. Opin. Neurobiol.* **4**, 909-914 (1994).
5. Nusbaum, M. P., Weimann, J. M., Golowasch, J. & Marder, E. *J. Neurosci.* **12**, 2706-2714 (1992).
6. Coleman, M. J. & Nusbaum, M. P. *J. Neurosci.* **14**, 6544-6552 (1994).
7. Selverston, A. I. & Moulins, M. *The Crustacean Stomatogastric System* (Springer, Berlin, 1987).
8. Harris-Warrick, R. M., Marder, E., Selverston, A. I. & Moulins, M. *Dynamic Biological Networks: The Stomatogastric Nervous System* (MIT Press, Massachusetts, 1992).
9. Weimann, J. M. & Marder, E. *Curr. Biol.* **4**, 896-902 (1994).
10. Weimann, J. M., Meyrand, P. & Marder, E. *J. Neurophysiol.* **65**, 111-122 (1991).
11. Harris-Warrick, R. M., Nagy, F. & Nusbaum, M. P. in *Dynamic Biological Networks: The Stomatogastric Nervous System* (eds Harris-Warrick, R. M., Marder, E., Selverston, A. I. & Moulins, M.) 87-137 (MIT Press, Massachusetts, 1992).
12. Coleman, M. J., Nusbaum, M. P., Courriel, I. & Claiborne, B. J. *J. comp. Neurol.* **325**, 581-594 (1992).
13. Norris, B. J., Coleman, M. J. & Nusbaum, M. P. *J. Neurophysiol.* **72**, 1451-1463 (1994).
14. Segev, I. *J. Neurophysiol.* **63**, 987-998 (1990).
15. Toth, P. T., Bindokas, V. P., Bleakman, D., Colmers, W. F. & Miller, R. J. *Nature* **364**, 635-639 (1993).
16. Cattaert, D., El Manira, A. & Clarac, F. *J. Neurophysiol.* **67**, 610-624 (1992).
17. Small, S. A., Cohen, T. E., Kandel, E. R. & Hawkins, R. D. *J. Neurosci.* **12**, 1616-1627 (1992).
18. Haydon, P. G., Man-Son-Hing, H., Doyle, R. T. & Zoran, M. J. *J. Neurosci.* **11**, 3851-3860 (1991).
19. Bennett, M. V. L. & Verselis, V. K. *Semin. Cell Biol.* **3**, 29-47 (1992).
20. Coleman, M. J. thesis, Univ. Alabama at Birmingham (1995).
21. Harris-Warrick, R. M. & Marder, E. *Rev. Neurosci.* **14**, 39-58 (1991).
22. Marder, E. & Weimann, J. M. in *Neurobiology of Motor Programme Selection: New Approaches to Mechanisms of Behavioral Choice* (eds Kien, J., McCrohan, C. & Winlow, W.) 3-19 (Manchester Univ. Press, Manchester, 1992).
23. Rossignol, S. & Dubuc, R. *Curr. Opin. Neurobiol.* **4**, 894-902 (1994).
24. Grillner, S. et al. *Trends Neurosci.* **18**, 270-279 (1995).
25. Spruston, N., Schiller, Y., Stuart, G. & Sakmann, B. *Science* **268**, 297-300 (1995).
26. Magee, J. C. & Johnston, D. *Science* **268**, 301-304 (1995).

27. Alford, S., Christenson, J. & Grillner, S. *Eur. J. Neurosci.* **3**, 107-117 (1991).
28. Sillar, K. T. & Simmers, A. J. *J. Neurosci.* **14**, 2636-2647 (1994).
29. Katz, P. S. & Frost, W. N. *J. Neurosci.* **15**, 6035-6045 (1995).
30. Dixon, D. B., Takahashi, K.-I. & Copenhagen, D. R. *Neuron* **11**, 267-277 (1993).

ACKNOWLEDGEMENTS. We thank D. M. Blitz, R. L. Calabrese, D. Combes, J. Raper and A. J. Simmers for useful discussions and for reading earlier versions of the manuscript. Supported by the NIH and NSF (M.P.N.), Foundation del DUCA (P.M.) and the HFSP.

## Altered *Hox* expression and segmental identity in *Mll*-mutant mice

Benjamin D. Yu, Jay L. Hess, Susan E. Horning, Gary A. J. Brown & Stanley J. Korsmeyer\*

Howard Hughes Medical Institute, Division of Molecular Oncology, Departments of Medicine and Pathology, Washington University School of Medicine, St Louis, Missouri 63110, USA

THE mixed-lineage leukaemia gene (*MLL/HRX/ALL-1*) is disrupted by chromosomal translocation in human acute leukaemias that often display mixed lymphoid-myeloid phenotypes and present in infancy<sup>1-4</sup>. *MLL* possesses a highly conserved SET domain also found in *Drosophila trithorax* (*trx*) and *Polycomb* group (*Pc-G*) genes, which are known to regulate homeotic genes (*HOM-C*) in a positive or negative fashion, respectively<sup>5</sup>. *Mll* was targeted in mice by homologous recombination in embryonic stem (ES) cells to assess its role in pattern development. *Mll* heterozygous (+/-) mice had retarded growth, displayed haematopoietic abnormalities, and demonstrated bidirectional homeotic transformations of the axial skeleton as well as sternal malformations. *Mll* deficiency (-/-) was embryonic lethal. Anterior boundaries of *Hoxa-7* and *Hoxc-9* expression were shifted posteriorly in *Mll* +/- embryos, but their expression was abolished in *Mll* -/- embryos. Thus *Mll* is required for proper segment identity in mammals, displays haplo-insufficiency, and positively regulates *Hox* gene expression.

We disrupted the murine *Mll* gene by inserting a *lacZ* reporter in-frame proximal to motifs conserved with *trx* and *Pc-G* genes (Fig. 1a). Loss of these motifs in *trx*<sup>B11</sup> is known to abolish *trithorax* activity in *Drosophila*<sup>6,7</sup>. Four independent, homologous recombinant D3 ES cell clones generated chimaeric males and transmitted the disrupted allele to the germ line (Fig. 1b). Reverse transcriptase-polymerase chain reaction (RT-PCR) confirmed the loss of *trx*-homologous motifs and the presence of 5' regions of *Mll* (Fig. 1c). Staining of heterozygous mice by  $\beta$ -galactosidase (Fig. 1d) and RNA *in situ* hybridization (not shown) indicated that *Mll* is widely expressed throughout embryonic development from the earliest tested time of embryonic day 7 (E7.0) into adulthood. Pronounced expression was noted in the embryonic nervous system, somites and somite derivatives. Therefore expression of *Mll* overlaps with known regions of *Hox* expression<sup>8</sup> (Fig. 1d, e).

*Mll* +/- mice demonstrated an overt phenotype and provided an unexpected opportunity to explore the role of *Mll*. Heterozygous mice were small at birth and had retarded growth (Table 1). Other consistent abnormalities include anaemia and decreased platelets, although morphology of haematopoietic cells was normal (Table 1). Maturation of all myeloid and lymphoid populations was observed; however, B-cell populations were consistently reduced in number in *Mll* +/- mice (not shown). *Mll* +/- females are hypofertile, yet five successful heterozygous matings resulted in 25 live pups, but none were *Mll* -/-, indicating that *Mll* deficiency is embryonic lethal. We recovered *Mll* -/- embryos at the expected frequency (+/+, 9; +/-, 18; -/-, 8) between E9.0 and E12.5. E10.5 was the latest point

\* To whom correspondence should be addressed.



when *Mll*  $-/-$  embryos were still viable according to the presence of cardiac contractions.

Segment abnormalities were present in *Mll*  $+/-$  mice, indicating disordered identity of cervical, thoracic and lumbar regions, and demonstrating the importance of *Mll* in pattern formation. Alizarin red-alcian blue staining indicated C2 and sternal malformations (Fig. 2a-c, Table 1). Anterior transformations of C7 to C6 identity and T3 to T2 identity were noted in *Mll*  $+/-$  skeletons (Fig. 2d-f, Table 1). *Mll*  $+/-$  newborns also had posterior transformations, with shifts of T13 to a L1 phenotype and L6 to S1 (Fig. 2g-i, Table 1). These transformations result in C7/T13/L5 or C7/T12/L6 configurations in contrast to the C7/T13/L6 wild-type configuration. Anterior or posterior homeotic transformations of the axial skeleton have been observed in *Hox*-deficient mice<sup>9</sup>. Although bidirectional transformations are rare in *Hox*-mutant mice<sup>10</sup>, they are characteristic of *trx* mutations which show transformation of both prothorax (T1) and metathorax (T3) towards mesothorax (T2) identity<sup>11,12</sup>.

*Hox* genes are important determinants of the mammalian body plan, and are expressed in a restricted anteroposterior pattern that coincides with their physical position in the cluster. Tight regulation of *Hox* expression boundaries and dosage are critical for specification of segment identity<sup>9</sup>. Whole-mount embryo hybridization and RNA *in situ* analysis detected caudal shifts in the anterior expression boundary of representative *Hox* genes in *Mll*  $+/-$  embryos. *Hoxc-9* mesodermal expression normally spans prevertebra (pv) 15 to pv27 segments at E10.5 (ref. 13) (Fig. 3a), but in *Mll*  $+/-$  embryos the anterior limit of *Hoxc-9* was shifted caudally 2-3 segments to pv17, pv18 (Fig. 3b). *Hoxa-7* regulation was also affected in *Mll*  $+/-$  embryos: its anterior limit of expression was pv11 (Fig. 3e) instead of the pv10 boundary in wild-type mice<sup>8</sup> (Fig. 3d). E11.5, as well as E10.5, *Mll*  $+/-$  embryos consistently demonstrated caudal shifts, and defects in *Hoxa-7* were detectable as early as E9.5 (not shown). RNA *in situ* analysis of *Mll*  $+/-$  embryos at E12.5 confirmed the defects in *Hoxa-7* expression in both mesodermal

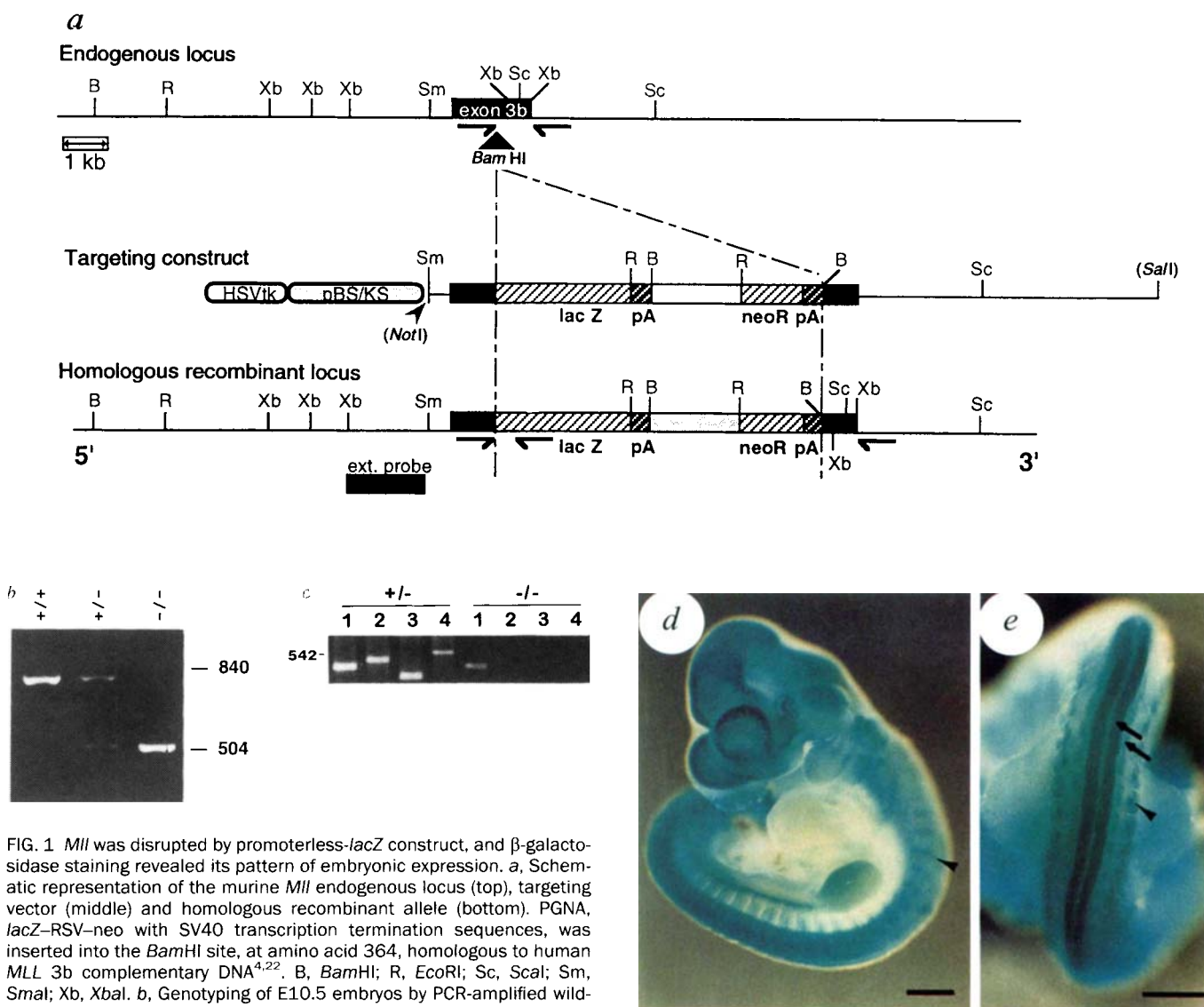


FIG. 1 *Mll* was disrupted by promoterless-*lacZ* construct, and  $\beta$ -galactosidase staining revealed its pattern of embryonic expression. a, Schematic representation of the murine *Mll* endogenous locus (top), targeting vector (middle) and homologous recombinant allele (bottom). PGNA, *lacZ*-RSV-neo with SV40 transcription termination sequences, was inserted into the *Bam*HI site, at amino acid 364, homologous to human *MLL* 3b complementary DNA<sup>4,22</sup>. B, *Bam*HI; R, *Eco*RI; Sc, *Sal*I; Sm, *Sma*I; Xb, *Xba*I. b, Genotyping of E10.5 embryos by PCR-amplified wild-type (840 bp) versus disrupted (504 bp) alleles. Distinct 3' primers for wild-type and disrupted alleles share a common 5' primer and are indicated by half-arrows in a. c, RT-PCR of AT-hook (lane 1), zinc-finger cluster 1 (lane 2), zinc-finger cluster 2 (lane 3), and C-terminal SET domain (lane 4) from *Mll*  $+/-$  and  $-/-$  E10.5 embryos. d, Lateral view of  $\beta$ -gal staining of E10.0  $+/-$  embryo, showing expression of *Mll* in neural and somitic tissue. Expression in somites was concentrated in a band (arrowhead) but was diffuse in posterior regions reflecting different stages of somite maturation. e, Dorsal view, forelimb level, showing that *Mll* was present throughout the neural tube, spinal ganglia (arrows),

sclerotome (arrowhead) and dermomyotome (confirmed in coronal sections, not shown). Scale bars, 0.5 mm.

METHODS. PGNA was cloned into a 7-kb *Sma*I-*Sal*I 129SV *Mll* genomic fragment, flanked by herpes simplex virus-thymidine kinase (HSV-tk). Primers were generated as previously described (ref. 22 and unpublished data). Generation of knockout mice, PGNA vector,  $\beta$ -gal staining, and staging of embryonic age have been described previously<sup>23-26</sup>. 1.5 kb *Sma*I-*Spe*I external or 0.9 kb *Sac*I internal probe hybridize to 3.5 kb *Xba*I wild-type and 10.5 kb recombinant fragments.

TABLE 1 Phenotype of *Mll* +/- mice

|                                                            | +/+ (n=43)  | +/- (n=33)  | P-value                |
|------------------------------------------------------------|-------------|-------------|------------------------|
| (a) Mass (g)                                               |             |             |                        |
| Newborn                                                    | 1.61 ± 0.34 | 1.42 ± 0.19 | 3.7 × 10 <sup>-4</sup> |
| 3-week                                                     | 12.7 ± 1.8  | 6.8 ± 0.8   | 8.8 × 10 <sup>-7</sup> |
| (b) Peripheral blood                                       |             |             |                        |
| White blood cells (×10 <sup>3</sup> )                      | 7.1 ± 3.4   | 6.8 ± 2.8   | 0.855                  |
| Red blood cells (×10 <sup>6</sup> cells μl <sup>-1</sup> ) | 9.7 ± 1.3   | 7.7 ± 0.7   | 0.010                  |
| Haemoglobin (g dl <sup>-1</sup> )                          | 15.7 ± 1.3  | 12.9 ± 1.1  | 0.002                  |
| Platelets (×10 <sup>3</sup> μl <sup>-1</sup> )             | 1,078 ± 145 | 822 ± 208   | 0.017                  |
| (c) Axial skeletal abnormalities*                          | +/+ (n=29)  | +/- (n=22)  |                        |
| C2 neural arch†                                            | 4           | 14          |                        |
| C6→C5                                                      | 1           | 7           |                        |
| C7→C6                                                      | 2           | 17          |                        |
| T2→T1                                                      | 2           | 16          |                        |
| T3→T2                                                      | 2           | 18          |                        |
| T13→L1                                                     | 1           | 8           |                        |
| L6→S1                                                      | 10          | 19          |                        |
| (d) Sternal abnormalities                                  | +/+ (n=53)  | +/- (n=35)  |                        |
| Manubriosternal‡                                           | 0           | 20          |                        |
| Xiphosternal§                                              | 6           | 17          |                        |

\* Both unilateral and bilateral abnormalities were included in the axial skeletal analysis.

† C2 abnormalities were predominantly neural arch defects, such as thickened or split neural arch defects and ectopic ossification between C1 and C2.

‡ Manubriosternal defects included asymmetries and fusions between manubrium and sternum.

§ Xiphosternal anomalies affected sternum 3–4 and the xiphoid process. Both *Mll* +/+ and *Mll* +/- demonstrated low-frequency asymmetry (~5%); *Mll* +/- exclusively showed fusions between sternum 3 and 4 and ossification occurring between sternum 4 and xiphoid process.

and neuronal compartments (Fig. 3g, h). Strikingly, *Hoxa-7* and *Hoxc-9* expression were undetectable in E10.5 *Mll* -/- embryos by whole-mount staining (Fig. 3c, f), despite the presence of normal sites of *Hox* expression, such as somites, spinal ganglia, limb buds and neural tube (not shown).

Thus *Mll* is required for the proper expression of *Hox* genes found in different clusters and paralogue groups. *Mll*-deficient embryos were embryonic lethal and failed to express representative *Hox* genes, whereas *Mll* heterozygotes possessed aberrations of segment identity and caudal shifts in the anterior *Hox* boundaries. Genetic mechanisms such as imprinting or gene dosage might account for a phenotype in heterozygous animals. However, there is no evidence for parental imprinting, as both maternally and paternally derived *Mll* alleles were expressed (not shown). *Mll*-mutant mice share many similarities with *trx* mutants. We know that *trx* homozygous mutants are also embryonic lethal and show defective *HOM-C* expression at stage 10–11 despite correct *HOM-C* boundaries one stage earlier<sup>14</sup>. Maintenance rather than initial specification of *HOM-C* boundaries requires *trx-G* and *Pc-G* genes<sup>15</sup>. The frequency of homeotic transformations and *Hox* expression defects appear more severe in *Mll* +/- than heterozygous deficiencies of *trx*. Motifs present in *Mll* but not *trx* may confer further transcriptional activities. Whether *Hox* expression is ever initiated in *Mll* -/- embryos will require an extensive analysis of more *Hox* genes at earlier time points. The disruption of *Mll* reported here would result in a truncated *Mll* that retains the AT-hook motifs. Because AT-hook motifs do not provide sequence-specific binding<sup>16</sup>, it is unlikely that the observed *Hox* expression abnormalities are neomorphic. Instead, as with *trx-G/Pc-G*, gene dosage appears to be important in the *Mll* +/- phenotype. Dosage sensitivity has also been demonstrated with overexpression of the mammalian *Bmi-1* oncogene, a *Pc-G* member, which showed a posterior shift in *Hoxc-5* expression<sup>17</sup>. The sensitivity of *Mll*, *trx-G* and *Pc-G* to dosage may represent rate-limiting or competitive inter-

actions. It has been suggested that *trx-G/Pc-G* regulate *Hox* expression at the level of chromatin organization, which is epigenetically inherited through multiple cell divisions<sup>15,18</sup>. Differences in segmentation between *Drosophila* and mammals, and in the kinetics of *HOM-C/Hox-C* expression in proliferating tissue, could also affect the severity of defects resulting from *Mll* versus *trx* deficiency<sup>19,20</sup>.

The generation of *Mll*-mutant mice establishes a role for *Mll* in the positive regulation of *Hox*, which bears similarity with *trx* rather than *Pc-G* genes. Involvement of mammalian *MLL* and *Bmi-1* in haematopoietic malignancy may also involve *Hox* dysregulation, as emerging data support a function for *Hox* genes in normal haematopoiesis<sup>21</sup>. These data suggest a model

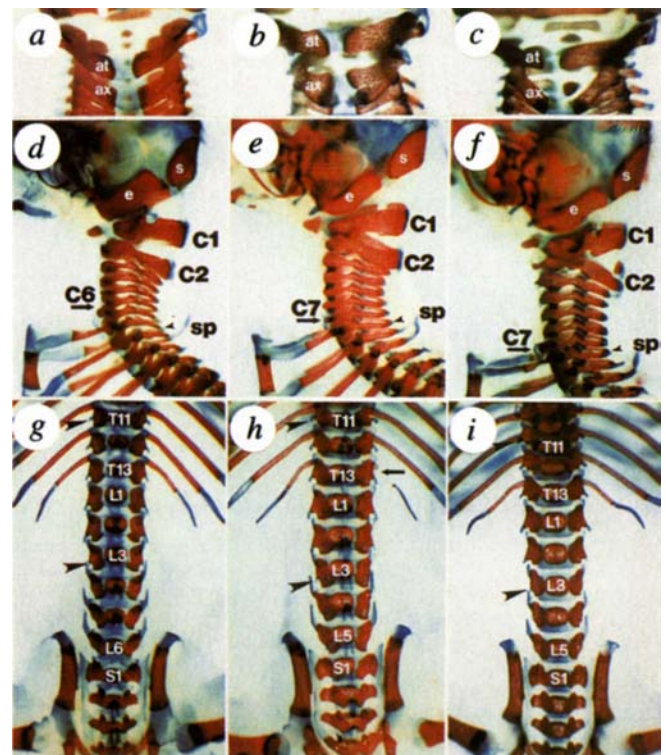


FIG. 2 Anterior and posterior transformations of axial skeletal identity in *Mll* +/- mice. Skeletons of newborns were prepared by alizarin red-alcian blue staining. C2 (ax) abnormalities such as splitting along neural arch (b) and ectopic points of ossification between C1 (at) and C2 (c) were observed in *Mll* +/- skeletons, but not in wild-type mice (a). *Mll* +/- newborns also displayed anterior transformations of C6, C7, T2 and T3. The tuberculum anterior on C6 (arrow) present in wild-type animals (d) was shifted to C7 in *Mll* +/- mice (e, f); s, supraoccipital bone; e, exoccipital bone. T1 is indicated by an arrowhead in lateral views of skeletal preparations (d-f). The exaggerated spinous process (sp), a characteristic of T2 in *Mll* +/- mice (d), was repeated on T3 (e) or shifted to T3 (f) in *Mll* +/- mice. Other indications of altered T2 identity included abnormal vertebral positioning of T2 rib to pre- or mid-manubrium rather than at the normal manubriosternal junction (not shown). T13, lumbar, and sacral identities were also effected in *Mll* +/- mice (g-i). T13 was commonly shifted to a L1 phenotype, either with complete loss of rib (not shown) or partial loss (h). *Mll* +/- newborns with and without T13 abnormalities showed five lumbar segments (h, i), while *Mll* +/- pups generally had six lumbar segments (g). L1, L2 and L3 were not remarkably altered in *Mll* +/- mice. The same pattern of T11-L3 caudal articular processes (arrowheads) was noted in *Mll* +/- (g) and *Mll* +/- (h, i) mice. The L6 identity appears to be transformed posteriorly to a S1 phenotype as the number of cervical and thoracic segments were not altered in *Mll* +/- mice.

METHODS. Newborns were generated from *Mll* +/- males bred to C57BL6/C3H F1 females and analysed by alizarin red-alcian blue staining as described previously<sup>27,28</sup>.



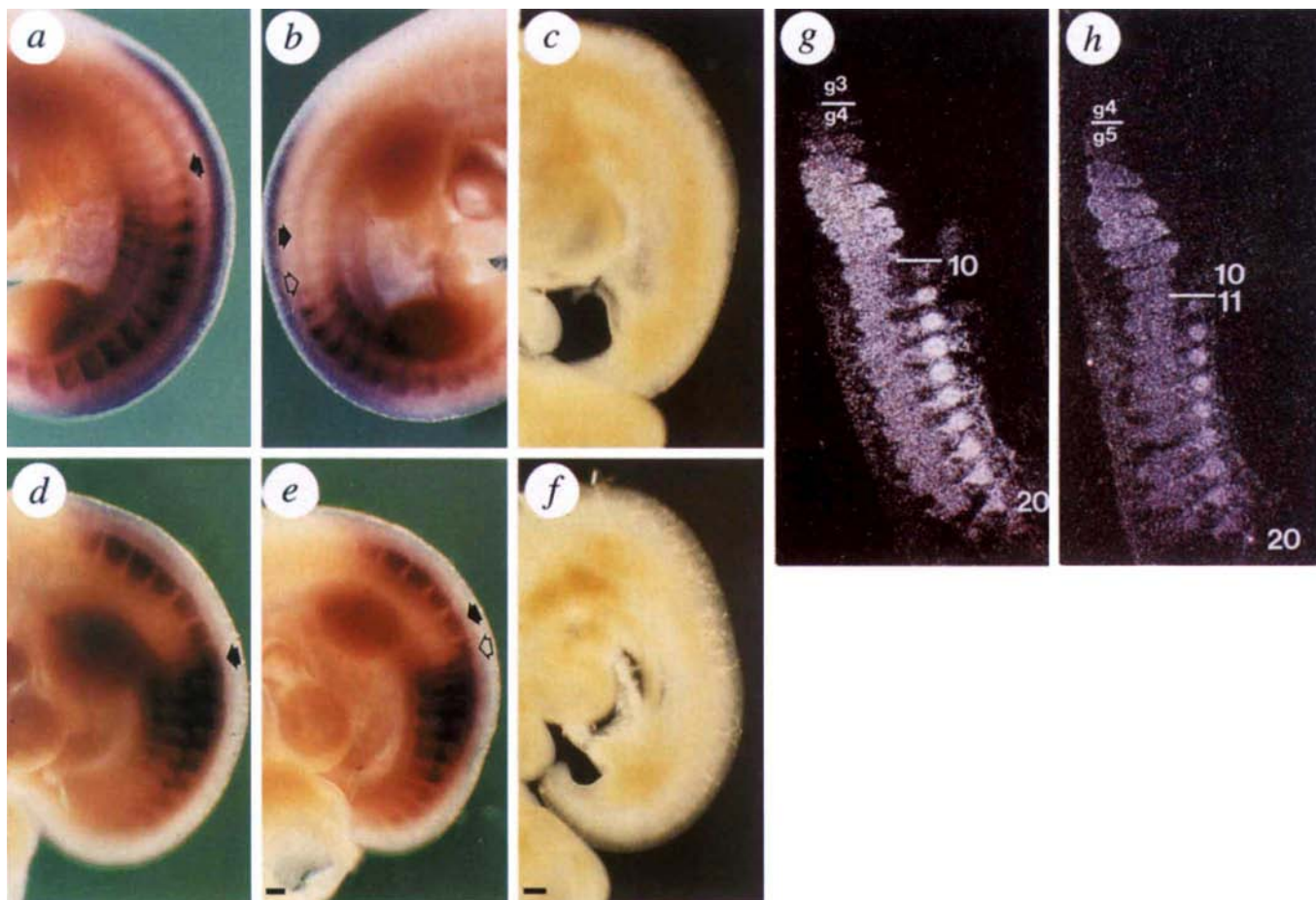


FIG. 3 Defective *Hox* regulation in *Mll*  $+/ -$  and  $-/-$  embryos. *Hoxc-9* (a–c) and *Hoxa-7* (d–f) whole-mount *in situ* hybridization of E10.5 embryos, and *Hoxa-7* RNA (g, h) *in situ* hybridization sections of E12.5 embryos. *Hoxc-9* expression was limited anteriorly at prevertebral level, pv15 in *Mll*  $+/ +$  embryos (a). This anterior boundary was shifted posteriorly 2–3 prevertebrae in *Mll*  $+/ -$  embryos (b). *Mll*  $-/-$  embryos were deficient in *Hoxc-9* expression (c). *Hoxa-7* showed a less severe defect than *Hoxc-9*. The wild-type *Hoxa-7* pattern spanned pv10 to pv20 (d), whereas the *Hoxa-7* anterior border in *Mll*  $+/ -$  embryos was shifted one segment caudally to pv11 (e). Like *Hoxc-9*, *Hoxa-7* expression was not detected in *Mll*  $-/-$  embryos (f). Normal (filled arrows) and defective (open arrows) anterior boundaries of *Hox* expression are labelled (a–f). RNA *in situ* section hybridization with a *Hoxa-7* probe demonstrated pv10–pv20 anteroposterior expression boundaries in *Mll*  $+/ +$  embryos

(g), in contrast to the pattern in *Mll*  $+/ -$  embryos, pv11–pv20 (h). Posterior boundary of *Hoxa-7*, pv20, in *Mll*  $+/ -$  embryos was unaffected. Anterior boundary of *Hoxa-7* expression in spinal ganglia was also shifted from the fourth spinal ganglion (g4) to the fifth (g5). White bars (g, h) indicate anterior limits of segmented mesodermal and neuronal expression of *Hoxa-7*. Developmental abnormalities of *Mll*  $-/-$  embryos will be described further elsewhere. METHODS. *In situ* hybridization protocols and patterns of *Hox* expression were described previously<sup>8,13,29</sup>. Antisense *Hoxc-9* (905-bp *Bam*HI) and *Hoxa-7* (363-bp *Sma*I) riboprobes were generated with Stratagene T3 RNA Polymerase from *Xba*I- or *Bam*HI-linearized templates, respectively. Sense probes for both *Hoxa-7* and *Hoxc-9* were used as controls and showed no hybridization.

in which the single event of *MLL* translocation results in two-hit oncogenesis. A *MLL* fusion product with a partner gene could confer gain-of-function activity, and simultaneous *MLL* haplo-insufficiency would contribute to the disordered cell fate of mixed-lineage leukaemias. □

Received 19 July; accepted 4 October 1995.

- Rowley, J. D. *Semin. Cancer Biol.* **4**, 377–385 (1993).
- Tkachuk, D. C., Kohler, S. & Cleary, M. L. *Cell* **71**, 691–700 (1992).
- Gu, Y. et al. *Cell* **71**, 701–708 (1992).
- Domer, P. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 7884–7888 (1993).
- Simon, J. *Curr. Opin. Cell Biol.* **7**, 376–385 (1995).
- Mazo, A. M., Huang, D.-H., Mozer, B. A. & David, I. B. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2112–2116 (1990).
- Breen, T. R. & Harte, P. J. *Mech. Dev.* **35**, 113–127 (1991).
- Knittel, T., Kessel, M., Kim, M. H. & Gruss, P. *Development* **121**, 1077–1088 (1995).
- Krumlauf, R. *Cell* **78**, 191–201 (1994).
- Jeannotte, L., Lemiux, M., Charron, J., Poirier, F. & Robertson, E. J. *Genes Dev.* **7**, 2085–2096 (1993).
- Ingham, P. & Whittle, R. *Molec. gen. Genet.* **179**, 607–614 (1980).
- Lindsley, D. L. & Zimm, G. G. *The Genome of Drosophila melanogaster* (Academic, San Diego, 1992).

- Erselius, J. R., Goulding, M. D. & Gruss, P. *Development* **110**, 629–642 (1990).
- Sedkov, Y., Tillib, S., Mizrokhi, L. & Mazo, A. *Development* **120**, 1907–1917 (1994).
- Singh, P. B. *J. Cell Sci.* **107**, 2653–2668 (1994).
- Reeves, R. & Nissen, M. S. *J. Biol. Chem.* **265**, 8573–8582 (1990).
- Alkema, M. J., van der Lugt, N. M. T., Bobeldijk, R. C., Berns, A. & van Lohuizen, M. *Nature* **374**, 724–727 (1995).
- Paro, R. *Trends Genet.* **11**, 295–297 (1995).
- Duboule, D. *Development (Suppl.)* 135–142 (1994).
- Tam, P. P. & Trainor, P. A. *Anat. Embryol.* **189**, 275–305 (1994).
- Lawrence, H. J. & Largman, C. *Blood* **80**, 2445–2453 (1992).
- Ma, Q. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 6350–6354 (1993).
- Mansouri, S. L., Thomas, K. R. & Capecchi, M. R. *Nature* **336**, 348–352 (1988).
- Le Mouellac, H., Lallemand, Y. & Brûlet, P. *Cell* **69**, 251–264 (1992).
- Hogan, B., Beddington, R., Constantini, F. & Lacy, E. *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory Press, New York, 1994).
- Kaufman, M. H. *The Atlas of Mouse Development* (Academic, San Diego, 1992).
- Kessel, M. & Gruss, P. *Cell* **67**, 89–104 (1991).
- Popesco, P., Rajtova, V. & Horak, J. *A Colour Atlas of Anatomy of Small Laboratory Animals* Vol. 2 (Wolfe, London, 1992).
- Wilkinson, D. G. & Nieto, M. A. *Meth. Enzym.* **225**, 361–373 (1993).

ACKNOWLEDGEMENTS. We thank P. Brûlet for PGNA vector, J. Merlie for *Hoxa-7* and *Hoxc-9* probes, and P. Dörm and S. Fakhrazadeh for materials and discussion. This work was supported by a grant from The Parker Hughes Trust, and B.D.Y. was supported by a training grant from the NIH.

# Myc but not Fos rescue of PDGF signalling block caused by kinase-inactive Src

M. Vittoria Barone\* & Sara A. Courtneidge\*†‡

\* Differentiation Programme, European Molecular Biology Laboratory, Postfach 10.2209, 69012 Heidelberg, Germany

† SUGEN Inc., 515 Galveston Drive, Redwood City, California 94063, USA

GROWTH factors such as platelet-derived growth factor (PDGF) elicit the transcriptional activation of a large number of immediate early genes (many of which encode transcription factors), and ultimately DNA synthesis<sup>1</sup>. Both AP1 and Myc are activated in fibroblasts in response to growth factor stimulation<sup>2-5</sup>, and various experiments suggest their importance in proliferation<sup>6-10</sup>. Src family kinases are required for PDGF (and other growth factors) to induce DNA synthesis<sup>11,12</sup>. We have examined which transcription factors, when constitutively expressed, 'rescue' the block elicited by dominant negative Src. We report here that Myc, but not Fos and/or Jun, was able to rescue the block. In contrast, Fos and Jun, but not Myc, rescued the block induced by dominant negative Ras. Our data suggest that Src kinases control the transcriptional activation of Myc.

The 'Ras pathway', leading to the transcriptional activation of genes such as *fos*, has recently been elucidated, and is necessary for PDGF-stimulated DNA synthesis<sup>13,14</sup>. Cells transformed with activated forms of Src have a constitutively activated Ras pathway<sup>15</sup> (initiated by the tyrosine phosphorylation of Shc<sup>16</sup>). During growth factor stimulation, Ras activation could be initiated by the tyrosine phosphorylation of Shc by the PDGF receptor itself, or by the associated Src family kinases<sup>16,17</sup>. Alternatively, recruitment of Grb2 to the membrane, through its association with the receptor-associated protein, Syp<sup>18</sup>, or the PDGF receptor itself<sup>19</sup> could initiate the activation. Given this apparent redundancy in the activation of Ras, we questioned whether Src family kinases were required for Ras activation. As a readout, we first examined Fos expression. One hour after PDGF stimulation, nuclear Fos staining was apparent, and in cells expressing a dominant negative form of Src (SrcK<sup>-</sup>), PDGF induction of Fos expression was unaffected (Fig. 1). We next tested whether constitutive expression of Fos could rescue the block caused by dominant negative forms of either Src or Ras. Cells were microinjected with plasmids encoding SrcK<sup>-</sup> alone, or SrcK<sup>-</sup>, Fos and Jun (Fig. 2a). Expression of SrcK<sup>-</sup> inhibited S phase entry (as measured by BrdU incorporation) by about 80%, as we have shown before<sup>11,12</sup>. No rescue of this block was achieved by constitutive expression of Fos and Jun, which were not by themselves inhibitory to the PDGF response (Fig. 2a). Fos and Jun also failed to rescue the SrcK<sup>-</sup> block to EGF- or CSF-1-mediated signalling (data not shown). Quiescent fibroblasts were also microinjected with a plasmid expressing dominant negative (N17)Ras alone, or N17Ras together with Fos and Jun plasmids. Fos and Jun (or Fos alone, data not shown) did rescue the block induced by N17Ras (Fig. 2b). These data suggest that Src family kinases are not absolutely required to activate the Ras pathway in response to PDGF, and further, that an important downstream component of the Ras pathway is the induction of Fos expression.

Some, but not all, investigators have detected elevated Myc levels in Src-transformed cells<sup>20,22</sup>, and Myc is an important component of some growth factor signalling<sup>8,10</sup>. We have previously shown that Src family kinases participate in CSF-1 receptor signalling<sup>12,23</sup>. A mutant CSF-1 receptor (Y809F) has been

described that fails to induce either Myc or growth in soft agar in response to ligand; forced expression of Myc from a heterologous promoter rescues the transformed phenotype<sup>24</sup>. The binding site for Src family kinases on the CSF-1 receptor has been mapped to Y561 and surrounding amino acids<sup>25</sup>. Nevertheless, one of the defects in the Y809F mutant is a reduced association and activation of Src family kinases<sup>23</sup>. This prompted us to test whether constitutive expression of Myc could rescue the block to DNA synthesis elicited by SrcK<sup>-</sup> (Fig. 3). SrcK<sup>-</sup> and Myc were coinjected into quiescent NIH-3T3 cells (for the PDGF and EGF experiments) and into NIH-3T3 cells engineered to express the CSF-1 receptor (for the CSF-1 experiments). In all cases, Myc was able to overcome the SrcK<sup>-</sup> arrest. Because

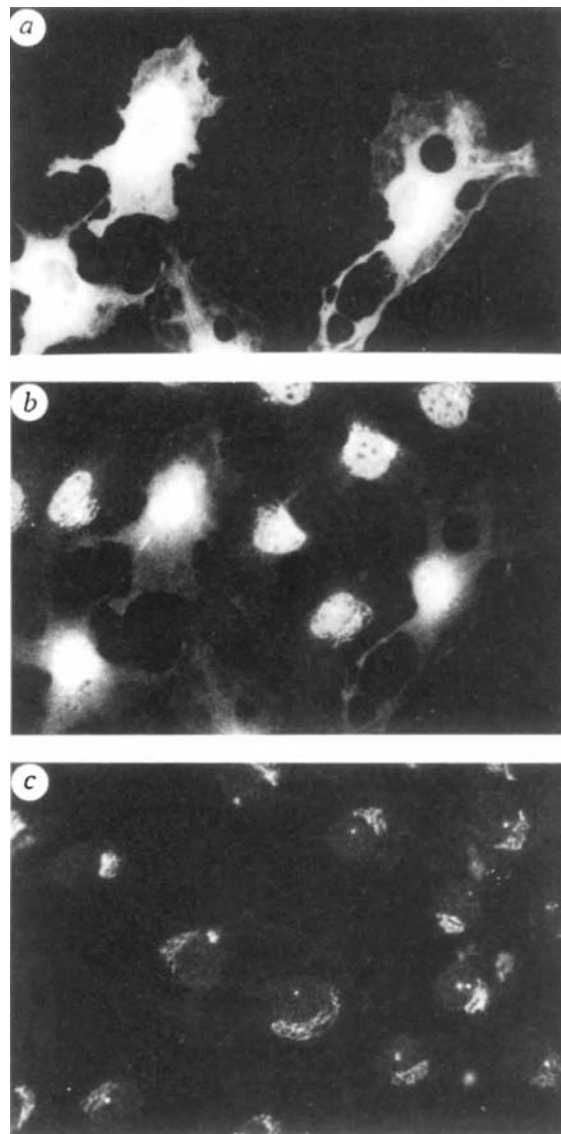


FIG. 1 SrcK<sup>-</sup> has no influence on Fos induction by PDGF. Arrested NIH-3T3 cells were either left quiescent (c) or injected with SrcK<sup>-</sup> (a and b), then 4–6 h later they were stimulated with PDGF for 1 h. a, Cells were stained with anti-cst.1 to detect the microinjected cells; b and c, cells stained for Fos.

METHODS. NIH-3T3 cells were processed as in Fig. 2. The coverslips were stained with a monoclonal mouse anti-Fos (1:200) (UBI) following the manufacturer's instructions, and with a secondary FITC-conjugated anti-mouse antibody (1:80) (UBI). Src was detected as in Fig. 2. The coverslips were analysed and photographed under fluorescence optics. Quantification from experiments such as these showed that more than 90% of the SrcK<sup>-</sup>-expressing cells were positive for Fos.

‡ To whom correspondence should be addressed.

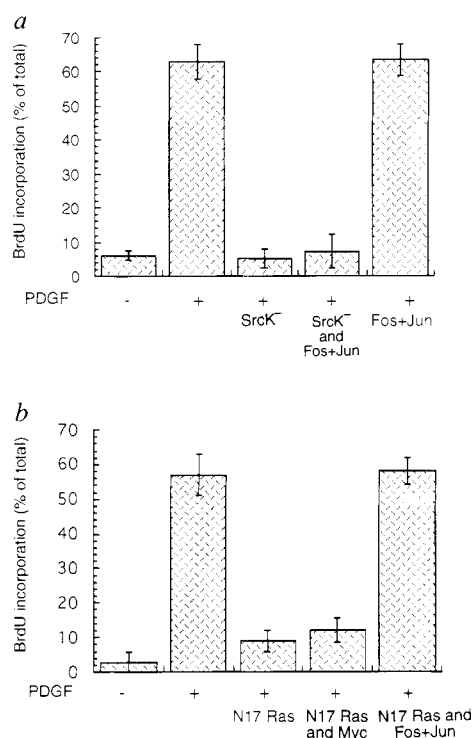


FIG. 2 The block to PDGF-induced S phase entry elicited by N17Ras, but not SrcK<sup>-</sup>, is overcome by Fos and Jun. **a**, Quiescent NIH-3T3 cells seeded on coverslips were microinjected with SrcK<sup>-</sup>, with and without Fos and Jun as shown, and either stimulated with PDGF or not stimulated. **b**, Quiescent NIH-3T3 cells seeded on coverslips were injected with N17Ras alone, or with N17Ras and Myc, or with N17Ras, Fos and Jun, then either stimulated with PDGF or not stimulated.

**METHODS.** NIH-3T3 cells were grown in Dulbecco's modified Eagle's medium (DMEM), 10% fetal calf serum (FCS), antibiotics and glutamine. Cells were then seeded on glass coverslips, arrested by culturing in DMEM, 0.5% FCS antibiotics, 1 mM glutamine, 5  $\mu\text{g ml}^{-1}$  transferrin, 5  $\mu\text{g ml}^{-1}$  insulin, for 24–48 h and then injected using an automated microinjection system (Zeiss) as described before<sup>28</sup>. The plasmids encoding SrcK<sup>-</sup> (50 ng ml<sup>-1</sup>), human cMyc (25 ng ml<sup>-1</sup>), N17Ras (50 ng ml<sup>-1</sup>), Fos (50 ng ml<sup>-1</sup>) and Jun (50 ng ml<sup>-1</sup>) were injected into the nucleus 6 h before adding PDGF and BrdU. The cells were fixed and processed for immunofluorescence 18 h later. Codetection of Src or N17Ras and BrdU was essentially as described before<sup>12</sup> using antibodies against Src ( $\alpha\text{cst.1}$ ) or a rat antibody for Ras (259). The extent of DNA synthesis was calculated by the formula: percentage of cells = (number of injected cells BrdU positive/total number of injected cells)  $\times$  100. For each lane more than 300 cells were analysed. Results from several experiments ( $n > 3$ ) have been averaged; the mean and the standard deviation of the mean are shown.

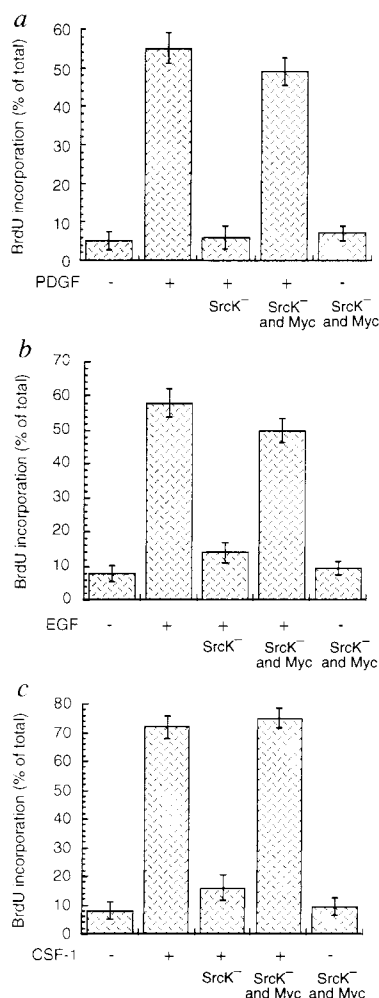


FIG. 3 The block to PDGF-, EGF- and CSF-1-stimulated S phase entry caused by SrcK<sup>-</sup> is overcome by Myc. **a**, Quiescent NIH-3T3 cells were injected with SrcK<sup>-</sup> or with SrcK<sup>-</sup> and Myc as shown, 6 h later stimulated with PDGF (10 ng ml<sup>-1</sup>), and 18 h later processed for immunofluorescence as in Fig. 2. **b**, Quiescent NIH-3T3 cells were injected with SrcK<sup>-</sup> or with SrcK<sup>-</sup> and Myc as shown, 6 h later stimulated with EGF (100 ng ml<sup>-1</sup>), and 18 h later processed for immunofluorescence as in Fig. 2. **c**, Quiescent NIH-3T3 cells expressing the CSF-1 receptor were microinjected with the constructs shown, 6 h later stimulated with CSF-1 (3,000 U ml<sup>-1</sup>), and 18 h later processed for immunofluorescence as in Fig. 2.



Fig. 4 Myc is required for PDGF-induced DNA synthesis, and its expression is regulated by Src family kinases. **a**, Quiescent NIH-3T3 cells were injected with plasmids encoding a dominant negative form of Myc (In 373) or SrcK<sup>-</sup> as shown, 6 h later stimulated with PDGF, and 18 h later processed for immunofluorescence as in Fig. 2. **b**, Quiescent NIH-3T3 cells were injected with either anti-cst.1 or control IgG as shown, stimulated with PDGF for 1.5 h, then lysed. RT-PCR was used to determine the amount of *myc* and GAPDH RNA. Shown is the ratio of *myc* to GAPDH.

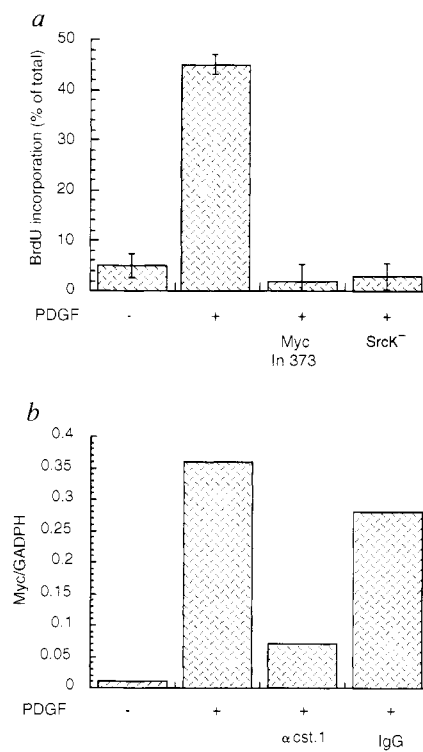
**METHODS.** Dominant negative Myc protocol. Quiescent NIH-3T3 cells were injected with a plasmid encoding a dominant negative form of Myc<sup>26</sup> or SrcK<sup>-</sup> at 50 ng ml<sup>-1</sup>. Six hours after injection, PDGF (25 ng ml<sup>-1</sup>) and BrdU were added to the cells. Eighteen hours later the cells were fixed and processed for immunofluorescence. Expression from the microinjected plasmids was detected with the 9E10 monoclonal antibody for Myc and anti-cst.1 for Src. The extent of DNA synthesis was calculated as in Fig. 2. In total, more than 200 cells were analysed. Results from 3 experiments have been averaged, and the mean and standard deviation of the mean are shown. RT-PCR protocol. Quiescent NIH-3T3 cells seeded on coverslips were injected with affinity purified anti-cst.1 or with a purified control IgG fraction. In each case, more than 70% of all the cells on the coverslip were injected (~2,000 cells). During the injection the temperature was held constant at 37 °C and the pH was maintained by adding 0.1 mM HEPES (pH 7.4) to the samples. Fifteen min after the injections were completed, PDGF was added to the samples, and 90 min later cells from each coverslip were lysed in 100 µl lysis buffer (0.5% Nonidet P-40, 10 mM Tris pH 8.5, 10 mM NaCl, 3 mM MgCl<sub>2</sub>). Reverse transcriptase reactions (RT Superscript Gibco-BRL) were done with 2 µl of each reaction mixture, following the manufacturer's instructions. PCR reactions were done in 100 µl, each reaction contained 25 µl of the RT reaction, 1 × PCR buffer (CETUS), 25 µM of each deoxynucleotide, 20 pmol of each 5' and 3' starter primer pair, two units of Taq DNA polymerase (CETUS). The PCR reaction consisted of 27 cycles at 94 °C × 1 min, 55 °C × 30 s, 72 °C × 1 min<sup>29</sup>. The oligonucleotides used were

|       |                             |
|-------|-----------------------------|
| Myc   | 5' CAAGAGGCGGACACACAACGTCT  |
|       | 3' AACTGTTCTCGTCGTTTCCTCAA  |
| GADPH | 5' CGGATGTCAACGGATTGGTCGTAT |
|       | 3' GGCCTTCTCCATGGTGGTGAAGAC |

Each PCR product (10 µl) was electrophoresed through 2% agarose gels, and hybridization was carried out as in ref. 30. Myc and GAPDH probes were obtained by end-labelling two oligos mapping in the internal part of each amplified band. Autoradiography was carried out overnight by using a Molecular Dynamics PhosphorImager and IMAGEQUANT software provided by the supplier. Quantitative analysis was with the GelAnalyzer Program.

at high concentration, Myc alone is sufficient to induce DNA synthesis<sup>2</sup>, it was important to show that the levels of Myc expression we achieved in these experiments did not render the cells growth factor independent (Fig. 3). In addition, the ability of Myc to rescue appeared to be specific to the Src pathway, because it failed to rescue the block elicited by dominant negative Ras (Fig. 2b).

Rescue of the SrcK<sup>-</sup> block by Myc predicts that Myc would be necessary for PDGF-induced DNA synthesis in NIH-3T3 cells. To test this, we microinjected a plasmid encoding a dominant negative form of Myc<sup>26</sup>, and measured BrdU incorporation in response to PDGF (Fig. 4a). Expression of this protein (Myc In373) completely inhibited the response to PDGF. Finally, it was not possible to examine Myc levels by immunofluorescence before and after growth factor stimulation, because of its low level of expression. As an alternative, we microinjected the majority (>70%) of the cells plated on a coverslip with either anti-cst.1 (an antibody that recognizes and inhibits Src, Fyn and Yes) or control IgG, lysed the cells and used reverse-transcribed polymerase chain reaction (RT-PCR) to determine the level of *myc* RNA, compared to a control (GAPDH RNA). We noted a marked reduction of *myc* RNA in the cells microinjected with the anti-cst.1, when compared to controls (Fig. 4b). Although an exact quantification by RT-PCR is not possible, these data



are consistent with our model that *myc* transcription is under the control of Src family kinases.

Our data strongly support the hypothesis that Src family kinases activate a pathway (most probably Ras independent) that culminates in the transcription of *myc*, and that is required for DNA synthesis in response to PDGF, EGF and CSF-1. We have termed this pathway the 'Src pathway'. Previous data, as well as experiments shown here, define a Ras pathway, which, among other things, results in activation of the transcription factor AP1 (ref. 27). We show here that constitutive expression of AP1, but not of Myc, is sufficient to overcome the inhibitory effects of N17Ras. Because both Src family kinases and Ras are required for S phase entry in response to PDGF, EGF and CSF-1, it follows that in normal cells not overexpressing activated components of these pathways, neither the Src pathway nor the Ras pathway would alone be sufficient to elicit DNA synthesis. Future experiments will address how activation of Src family kinases results in the production of Myc. □

Received 27 July; accepted 25 September 1995.

1. Cantley, L. C. et al. *Cell* **64**, 281–302 (1991).
2. Armelin, H. A. et al. *Nature* **310**, 655–660 (1984).
3. Greenberg, M. E. & Zitt, E. B. *Nature* **311**, 433–488 (1984).
4. Kruijer, W., Cooper, J. A., Hunter, T. & Verma, I. M. *Nature* **312**, 711–716 (1984).
5. Müller, R., Bravo, R., Burckhardt, J. & Curran, T. *Nature* **312**, 716–720 (1984).
6. Kelly, K., Cochran, B. H., Stiles, C. D. & Leder, P. *Cell* **35**, 603–610 (1983).

7. Holt, J. T., Gopal, T. V., Moulton, A. D. & Nienhuis, A. W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4794–4798 (1986).
8. Nishikura, K. & Murray, J. M. *Molec. cell. Biol.* **7**, 639–649 (1987).
9. Biro, S., Fu, Y.-M., Yu, Z.-X. & Epstein, S. E. *Proc. natn. Acad. Sci. U.S.A.* **90**, 654–658 (1993).
10. Wickstrom, E. L. et al. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1028–1032 (1988).
11. Twamley-Stein, G., Pepperkok, R., Ansong, W. & Courtneidge, S. A. *Proc. natn. Acad. Sci. U.S.A.* **90**, 7696–7700 (1993).
12. Roche, S., Kögl, M., Barone, M. V., Roussel, M. F. & Courtneidge, S. A. *Molec. cell. Biol.* **15**, 1102–1109 (1995).
13. Mulcahy, L. S., Smith, M. R. & Stacey, D. W. *Nature* **313**, 241–243 (1985).
14. Marshall, C. J. *Curr. Genet. Dev.* **4**, 82–89 (1994).
15. Smith, M. R., DeGudicibus, S. J. & Stacey, D. W. *Nature* **320**, 540–543 (1986).
16. McGlade, J., Cheng, A., Pelicci, G., Pelicci, P. G. & Pawson, T. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8869–8873 (1992).
17. Pelicci, G. et al. *Cell* **70**, 93–104 (1992).
18. Li, W. et al. *Molec. cell. Biol.* **14**, 509–517 (1994).
19. Arvidsson, A.-K. et al. *Molec. cell. Biol.* **14**, 6715–6726 (1994).
20. Sovova, V., Friis, R., Fidlerova, H. & Hlozaneck, I. *Int. J. Cancer* **53**, 983–987 (1993).
21. Kuchino, Y., Nemoto, K., Kawai, S. & Nishimura, S. *Jap. J. Cancer Res.* **76**, 75–78 (1985).
22. Welham, M. J., Wyke, J. A., Lang, A. & Wyke, A. W. *Oncogene* **5**, 161–169 (1990).
23. Courtneidge, S. A. et al. *EMBO J.* **12**, 943–950 (1993).
24. Roussel, M. F., Cleveland, J. L., Shurtleff, S. A. & Sherr, C. J. *Nature* **353**, 361–363 (1991).
25. Alonso, G., Koegl, M., Mazurenko, N. & Courtneidge, S. A. *J. biol. Chem.* **270**, 9840–9848 (1995).
26. Sawyers, C. L., Callahan, W. & Witte, O. N. *Cell* **70**, 901–910 (1992).
27. Stacey, D. W., Watson, T., Kung, H. & Curran, T. *Molec. cell. Biol.* **7**, 523–527 (1987).
28. Barone, M. V., Crozat, A., Tabaei, A., Philippon, L. & Ron, D. *Genes Dev.* **8**, 453–464 (1994).
29. Wong, H., Anderson, W. D., Cheng, T. & Rjabowol, K. T. *Analyt. Biochem.* **223**, 251–258 (1994).
30. Barone, M. V., Henchcliffe, F. E., Baralle, F. E. & Paoletti, G. *EMBO J.* **8**, 1079–1085 (1989).

ACKNOWLEDGEMENTS. We thank D. Bohmann for Jun and Fos expressing plasmids and for useful advice and discussions, D. Afar for In373 Myc, M. Roussel for CSF-1R-expressing cells, T. Ramqvist for the Myc-expressing plasmid, M. Jones for technical assistance, and G. Paoletti for help with the GelAnalyzer Program.

## Crystal structure of the annexin XII hexamer and implications for bilayer insertion

Hartmut Luecke, Bryna T. Chang,  
William S. Mailliard\*, David D. Schlaepfer\*  
& Harry T. Haigler\*

Stanford Synchrotron Radiation Laboratory, MS 69, Stanford,  
California 94305, USA

\* Department of Physiology and Biophysics, University of California,  
Irvine, California 92717, USA

ANNEXINS are a family of calcium- and phospholipid-binding proteins<sup>1,2</sup> implicated in a number of biological processes including membrane fusion<sup>3</sup> and ion channel formation<sup>4–7</sup>. The crystal structure of the annexin XII hexamer, refined at 2.8 Å resolution, forms a concave disk with 3–2 symmetry, about 100 Å in diameter and 70 Å thick with a central hydrophilic pore. Six intermolecular Ca<sup>2+</sup> ions are involved in hexamer formation. An additional 18 Ca<sup>2+</sup> ions are located on the perimeter of the disk, accessible only from the side of the hexameric disk. On the basis of the hexamer structure we propose here a new mode of protein–phospholipid bilayer interaction that is distinct from the hydrophobic insertion of typical membrane proteins. This speculative model postulates the Ca<sup>2+</sup>-dependent insertion of the hydrophilic annexin XII hexamer into phospholipid bilayers with local reorientation of the bilayer phospholipids.

We determined the structure of the hydra annexin XII (ref. 8) hexamer by the molecular replacement method (Fig. 1). The refined 2.8 Å resolution atomic structure reveals a 210K homohexamer with 3–2 symmetry in the asymmetric unit. The hexamer forms a concave disk about 100 Å in diameter and 70 Å thick with a prominent central hydrophilic pore situated along the local 3-fold axis. The individual annexin XII trimers, which are joined face-to-face in the hexamer, are similar to the annexin V trimers reported to form on phospholipid monolayers in two-

TABLE 1 Buried surface area and number of intermolecular salt bridges for various oligomers of annexin XII

| Aggregate           | Accessible surface area (Å <sup>2</sup> ) | Buried Surface area (Å <sup>2</sup> ) | Intermolecular salt bridges |
|---------------------|-------------------------------------------|---------------------------------------|-----------------------------|
| Monomer             | 13,643                                    | 0                                     | 0.0%                        |
| Dimer within trimer | 25,974                                    | 1,312                                 | 4.8%                        |
| Dimer across trimer | 24,800                                    | 2,486                                 | 9.1%                        |
| Trimer              | 36,997                                    | 3,932                                 | 9.6%                        |
| Hexamer             | 66,278                                    | 15,580                                | 19.0%                       |

The intermolecular interactions for the hexamer are extensive (19% of the accessible surface area per monomer) with a predominantly hydrophilic character. The dimer forming the intermolecular Ca<sup>2+</sup> site ('dimer across trimer') has considerably more buried surface area than the dimer within a trimer. Thus the 'dimer across trimers' may be the dimer observed in previous solution studies of annexins<sup>15–18</sup>. Other studies showed that annexin V forms trimers on planar monolayers<sup>9</sup> which are very similar to the trimers composing the annexin XII hexamer. The key interactions for annexin hexamer formation may be a combination of the interactions that caused the previously observed formation of dimers and trimers. **Methods.** Annexin XII was expressed in recombinant bacteria and purified as previously described<sup>9</sup>. Crystals (0.05 mm × 0.15 mm × 0.15 mm) were obtained by the hanging drop vapour diffusion method from a solution containing 7 mg ml<sup>-1</sup> annexin XII, 6% w/v PEG 6000 and 10 mM CaCl<sub>2</sub> at pH 7.8; X-ray diffraction data to 2.8 Å resolution were collected at 8 °C from one crystal (space group P2<sub>1</sub>, a = 71.2 Å, b = 179.2 Å, c = 100.7 Å, β = 97.6°) at beam line 7-1 at the Stanford Synchrotron Radiation Laboratory and reduced with the program MOSFLM<sup>27</sup>, yielding a completeness of 86.9% with an R<sub>MERGE</sub>(h) of 6.8%. The structure was solved by the molecular replacement method, using the human annexin I monomer without calcium<sup>11</sup> as a model (44% sequence identity). The self-rotation function (program GLRF<sup>28</sup>) suggested a local 3-fold axis nearly parallel to the crystallographic a axis. Visual inspection (program CHAIN<sup>29</sup>) of C $\alpha$ -traces oriented according to the 20 highest crossrotation solutions obtained with GLRF followed by translational searches with the program X-PLOR<sup>30</sup> yielded solutions consistent with a trimeric arrangement. The position of the second trimer was found by using the first trimer as the search model. The structure was refined with X-PLOR using 6-fold non-crystallographic symmetry (NCS) restraints. NCS B-factor restraints were kept at 2.0 Å<sup>2</sup> throughout refinement. NCS coordinate restraints were initially 300 kcal mol<sup>-1</sup> Å<sup>-2</sup> and gradually relaxed to 50 kcal mol<sup>-1</sup> Å<sup>-2</sup> to achieve the largest decrease in R<sub>FREE</sub> during refinement. At this point, the R<sub>WORK</sub> was 23.6% and the R<sub>FREE</sub> 27.8%. Subsequently, all observed structure factors with F > σ<sub>F</sub> (52,044 reflections between 10 Å and 2.8 Å) were used in refinement. The final crystallographic R-factor for the hexamer, consisting of 14,772 non-hydrogen protein atoms and 24 bound calcium ions, is 21.6% with an overall B-factor of 11.0 Å<sup>2</sup>. No solvent molecules have been included in the model. The root-mean-square deviations from ideal bond lengths and angles are 0.018 Å and 3.1°, respectively. All residues are in the allowed regions of the Ramachandran plot (PROCHECK<sup>27</sup>). The r.m.s. coordinate difference for main chain atoms between annexin XII monomers and annexin I<sup>9</sup> is 1.1 Å. The r.m.s. coordinate difference for main chain atoms between annexin XII monomers and annexin V<sup>8</sup> is 1.9 Å; this larger value is mainly due to the different conformation of the type II loop in domain III of annexin V. The atomic coordinates have been deposited with the Protein Data Bank (Brookhaven).

\* including Ca<sup>2+</sup>.

dimensional crystals<sup>9</sup>. The hexamer is stabilized by a number of specific interactions, including six intermolecular Ca<sup>2+</sup>-binding sites (Fig. 2), 18 intermolecular salt bridges as well as hydrogen bonds. In addition, 19% of the accessible surface area of each monomer is buried upon hexamer formation (Table 1). The convex faces of all six annexin XII monomers are buried at the trimer–trimer interface of the hexamer (Fig. 1c). This arrangement initially was surprising because the convex face is thought to be involved in lipid bilayer binding<sup>9,10</sup>. However, closer inspection reveals that 18 type II Ca<sup>2+</sup> sites of domains I, III and IV (see ref. 11 for numbering convention) of each of the six monomers are occupied by Ca<sup>2+</sup> and that they remain available for interaction with phospholipid head groups at the perimeter of the hexameric disk (Fig. 1a and d).

The six intersubunit Ca<sup>2+</sup> sites are involved in hexamer formation. They are formed by peptide carbonyls 68, 71 and the monodentate carboxyl of Glu 76 from one molecule and the monodentate carboxyl of Glu 105 from another molecule (annexin XII numbering convention, Fig. 2). Glu 105 is fully conserved in all reported annexin sequences<sup>1</sup>. The second carboxyl oxygen of Glu 76 also accepts a hydrogen bond from the peptide amide of residue 103 of the other molecule. Glu 76 also is conserved in other annexins. In addition, the carboxyl moiety of Asp 230 from the same molecule that provides Glu 105 is

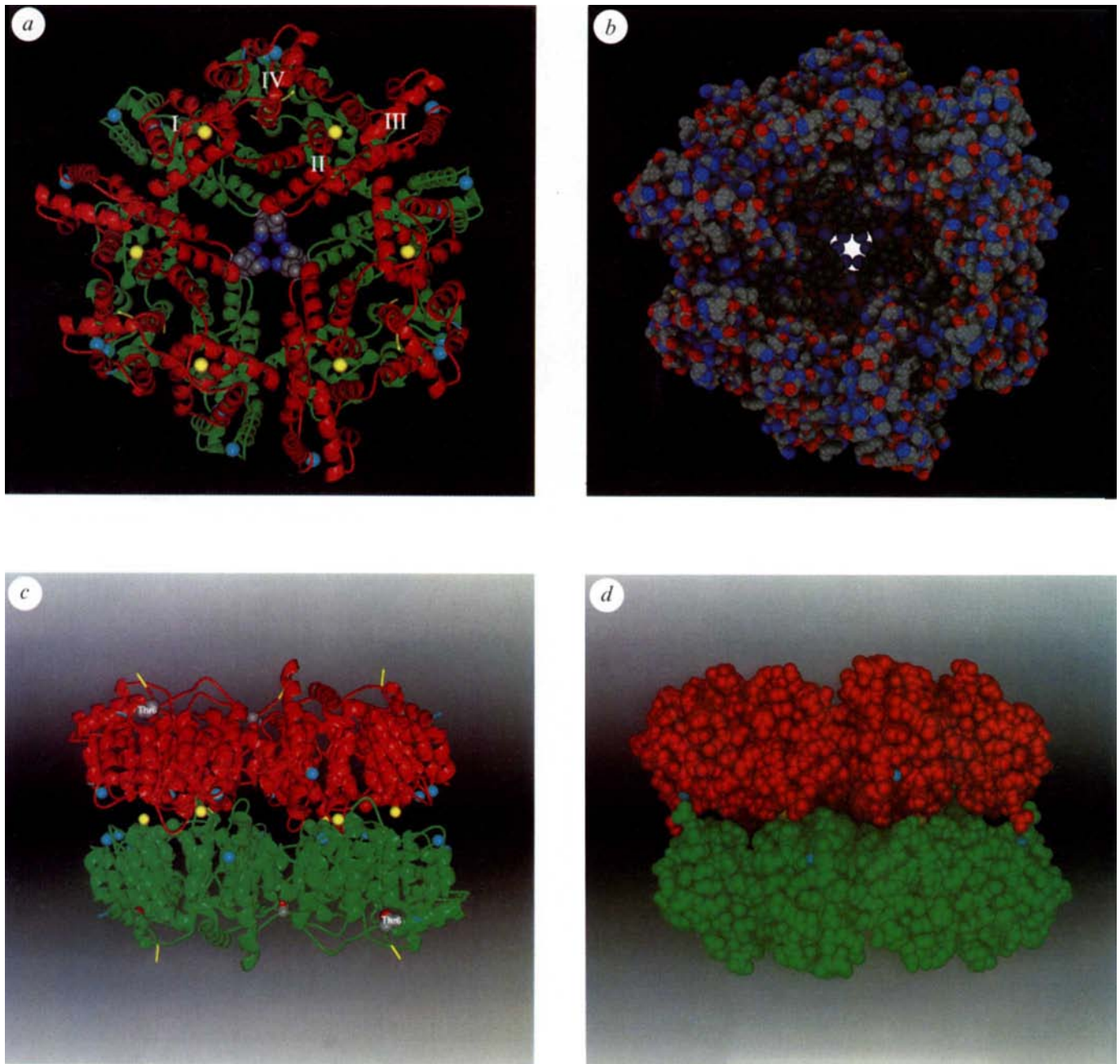


FIG. 1 *a*, Ribbon drawing of the annexin XII hexamer viewed down the local 3-fold axis. The upper trimer (red) and the lower trimer (green) assemble face-to-face. The individual domains of one molecule of the red trimer are marked with Roman numerals. The two types of  $\text{Ca}^{2+}$  sites are coloured differently: light-blue spheres represent the 18  $\text{Ca}^{2+}$  atoms bound to the type II sites of domains I, III and IV, situated at the perimeter of the disk; yellow spheres highlight the six  $\text{Ca}^{2+}$  ions bound to the intersubunit sites responsible for hexamer formation. The side chains of six equivalent lysine residues (Lys 132) are shown (space-filling model) in the centre of the hydrophilic pore. These lysines are solvent-accessible and the electron densities of their side chains are not well defined. In the refined model  $\epsilon$ -amino groups of the three Lys 132 within one trimer are 7 Å apart, between trimers they are 6 Å apart, defining the most constricted portion of the pore which is about 8 Å in diameter. The N termini of the individual monomers start near domain IV and are coloured yellow, the nearby C termini are coloured blue. *b*, Space-filling model of the annexin XII hexamer viewed down the local 3-fold axis, in the same orientation as *a*. The central hydrophilic pore is funnel-shaped, with most of the funnel surface being defined by helix IIC and the loop following it. The smallest diameter of the pore

is defined by six lysine side chains (Lys 132). In the refined structure these lysine side chains are not fully extended towards the centre of the pore, conceivably because of electrostatic repulsion. *c*, Hexamer viewed side-on along a local 2-fold axis. All 24  $\text{Ca}^{2+}$  sites are at or near the trimer-trimer interface. The 18 occupied type II sites (blue) are accessible at the perimeter of the disk. The domain II type II loops which are occupied by lysine  $\epsilon$ -amino groups from another molecule protrude across the central plane into the other trimer. Also shown are the six threonine (Thr 6) side chains at the faces of the hexamer which are targets for phosphorylation by protein kinase  $\text{C}^8$ . The threonine hydroxyls are pointing away from the bulk solvent, hydrogen bonding to the peptide carbonyl of residue 313. A phosphorylated Thr 6 would have to adopt a different conformation, most likely exposing the charged side chain to the bulk solvent, possibly causing a displacement of the centre N-terminal domain. *d*, Space-filling view along a local 2-fold axis, in the same orientation as *c*. Four of the 18 exposed type II  $\text{Ca}^{2+}$  ions (blue) are clearly visible, two more are barely visible at the sides. These sites remain accessible for interaction with phospholipid head groups. Also shown in yellow are two of the intermolecular  $\text{Ca}^{2+}$  ions.



FIG. 2 Atomic model in the vicinity of the intermolecular  $\text{Ca}^{2+}$  site. The  $\text{Ca}^{2+}$  has four proteinaceous ligands: backbone carbonyl oxygens of residues 68 and 71, and monodentate Glu 76 from the lower molecule (shaded green); the upper molecule (shaded red) coordinates with Glu 105. Glu 105 is conserved in all published annexin sequences. The average oxygen–calcium distance is 2.2 Å. In addition, the carboxyl moiety of Asp 230 is only 3.6 Å from the intermolecular  $\text{Ca}^{2+}$ . The second carboxyl oxygen of Glu 76 also accepts an intermolecular hydrogen bond from the peptide amide of residue 103. On the right-hand side the type II  $\text{Ca}^{2+}$ -binding loop of domain II is shown with Leu 101 penetrating a hydrophobic region between helices 1E and 1A of the lower molecule. The  $\text{Ca}^{2+}$  ion in this loop has been replaced by the  $\epsilon$ -amino group of Lys 68 of the lower molecule which is interacting with three peptide carbonyls of the loop. The side chain of Glu 142, which in the presence of  $\text{Ca}^{2+}$  acts as a capping residue, is now displaced away from the loop. Molecular graphics images were produced using the MidasPlus program from the Computer Graphics Laboratory, University of California, San Francisco (supported by the NIH).

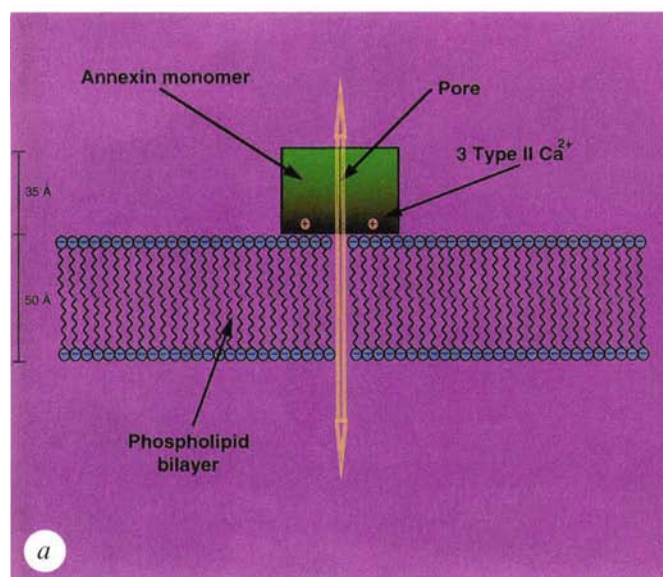
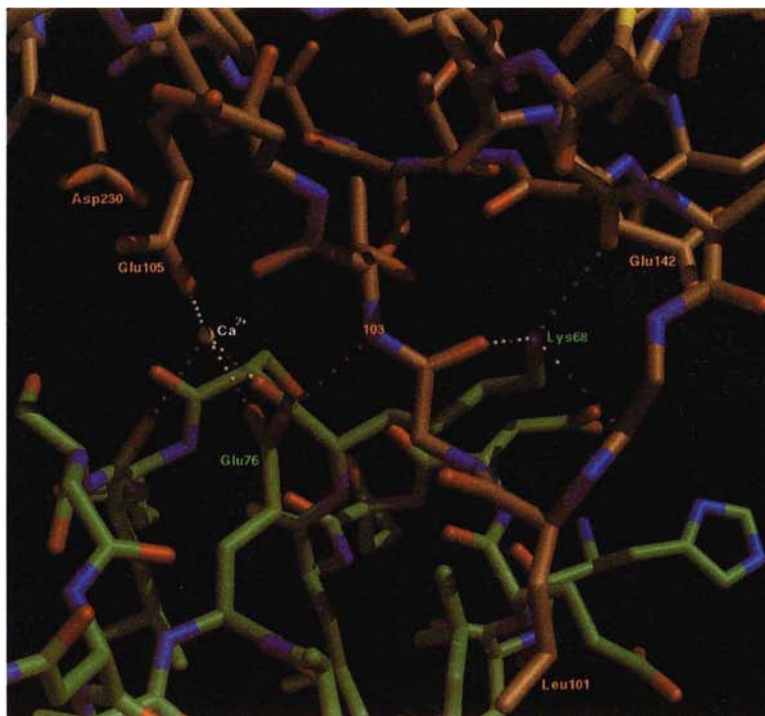
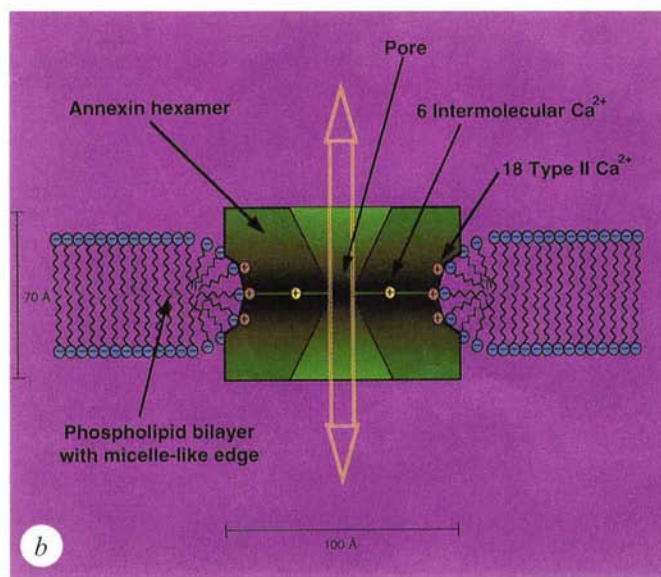


FIG. 3 Models for  $\text{Ca}^{2+}$ -dependent interaction between annexins and phospholipid bilayers. *a*, Peripheral binding of an annexin monomer to a planar bilayer (see ref. 7). The interaction is stabilized by three to six  $\text{Ca}^{2+}$  ions (shown in red) interacting with phospholipid head groups. The annexin monomer does not penetrate the hydrophobic core of the bilayer. The putative channel is formed by a very narrow hydrophilic region within the monomer. Ions have been proposed to cross the adjacent bilayer through transient or permanent water-filled pores<sup>6,7</sup>. However, theoretical diffusion-limited calculations indicate that this hydrophilic region is much too narrow to support the observed conductance. The maximal ion flux through a cylindrical pore of length  $l$  and radius  $a$  can be approximated by assuming diffusion-limited movement of ions<sup>24</sup>. The experimentally observed single-channel current for annexin V (2.8 pA for 50 mM  $\text{CaCl}_2$  at 90 mV, ref. 22) would require a substantially wider pore diameter than is present in the monomer. The central hydrophilic region in the annexin V monomer is about 15 Å in length and less than 1 Å in diameter<sup>25</sup>. Our calculations (see pages 296–298, equations (11–1) and (11–6) of ref. 24) indicate that a pore 15 Å in length would have to be at least 15 Å in diameter to produce single-channel currents of this magnitude. On the other hand, it should be noted that site-directed mutations in this region alter the channel properties of annexin V and support the proposal that this region mediates ion flux<sup>6,7</sup>. However, it is not possible to draw definitive conclusions



from this study because other mutations on the surface of annexin V distant from this region also modify the conductance pattern<sup>6,7</sup>. *b*, Annexin hexamer embedded in a bilayer by hydrophilic interaction. Although no direct data are available, we speculate that a local micelle-like edge of the bilayer allows the 18 type II affinity  $\text{Ca}^{2+}$  sites (shown in red) on the perimeter of the hexamer to interact with phospholipid head groups. This arrangement is symmetrical with respect to the bilayer and the hydrophilic funnel-shaped pore at the centre of the hexamer forms a putative pathway for ions to cross the bilayer. The charged side chains of the lysine residues at the mouth of the pore are likely to move in response to an external electric field and could thus be involved in voltage-gating. It should be noted that the very narrow hydrophilic regions within the monomers are lined up head-to-head in the hexamer. However, ion conductance through this narrow pathway in the hexamer is even less likely than through a single monomer (see *a*) because of increased path length. The hexamer is stabilized by 6 intersubunit  $\text{Ca}^{2+}$  ions (shown in yellow). The orientation of the hexamer is similar to the orientation in Fig. 1*d*. Membrane fusion could be helped in such an arrangement by the partial exposure of hydrophobic lipid tails at the micelle-like edge. Long-range hydrophobic interactions have been shown to be responsible for spontaneous bilayer fusion in model systems<sup>26</sup>.

situated 3.9 Å from the intermolecular  $\text{Ca}^{2+}$ , further stabilizing the complex. These intermolecular  $\text{Ca}^{2+}$  sites share three coordination sites with the intramolecular 'lanthanum' site 'Ca5' first reported for annexin V (ref. 10). Annexin I also showed  $\text{Ca}^{2+}$  binding at the type III site of domain I (ref. 11). These may be the low-affinity  $\text{Ca}^{2+}$  sites previously reported for annexin VII (ref. 12) and annexin II (ref. 13).

A further very specific intermolecular interaction involves the type II  $\text{Ca}^{2+}$ -binding sites of domain II. These sites are occupied by  $\text{Ca}^{2+}$  in all previously reported annexin monomer structures<sup>2,10,11</sup>. Interestingly, in the annexin XII hexamer these sites contain no bound  $\text{Ca}^{2+}$ , but instead they are occupied by the  $\epsilon$ -amino group of Lys 68 from a molecule of the opposing trimer, displacing the capping carboxylate moiety of Glu 142 (Fig. 2). Lysine  $\epsilon$ -amino groups replacing  $\text{Ca}^{2+}$  have been observed in phospholipase  $\text{A}_2$ <sup>14</sup> whose  $\text{Ca}^{2+}$  site has a topography very similar to the type II site of the annexins. Lys 68 is conserved in all published annexin sequences.

Annexin XII hexamer formation is not confined to crystals. Subsequent to the crystal structure determination, chemical crosslinking (W.S.M., H.L. and H.T.H., manuscript in preparation) and small angle X-ray scattering (H. T. Tsuruta, H.T.H. and H.L., manuscript in preparation) studies of annexin XII in solution demonstrated reversible  $\text{Ca}^{2+}$ -induced hexamer formation.

Although none of the previous crystallographic studies of annexins have revealed hexameric complexes<sup>2</sup>, we propose that other annexins are likely to form hexamers similar to the annexin XII hexamer. As noted above, several of the key residues involved in annexin XII hexamer formation are highly conserved in other annexins. This hypothesis also is supported by solution studies that report self-aggregation of annexins I (ref. 15), II (ref. 15), IV (refs 15, 16), V (refs 15, 17, 18), VI (refs 15, 16) and VII (refs 16, 19). Furthermore, electron microscopy studies of annexin V bound to phospholipid monolayers<sup>7,20</sup> revealed trimers with a structure similar to the trimers within the annexin XII hexamer. It is possible that hexamers were not detected in the electron microscopy studies of annexin V because monolayers were used rather than bilayers (see model in Fig. 3b).

The current model for annexin-phospholipid bilayer interaction is based on the crystal structures of annexin monomers<sup>2,10,11,21</sup> (Fig. 3a and ref. 7). The model proposes peripheral attachment of annexin monomers to the bilayer, mediated by interactions of the bound  $\text{Ca}^{2+}$  on the convex face of the monomer with the polar phospholipid head groups. This mode of interaction explains the reversible  $\text{Ca}^{2+}$ -dependent binding of annexins to the surface of bilayers, but does not offer a convincing mechanism to explain the perturbation of bilayer structure which causes membrane fusion or ion channel activity

observed for certain annexins *in vitro*. In particular, it is not clear how an annexin monomer bound peripherally on one side of the bilayer leaflet could effect symmetric (non-rectifying) ion flux across the bilayer<sup>22</sup>. In addition, the diameter of the proposed pore is too small to be reasonably expected to conduct the observed ion currents (see Fig. 3a).

On the basis of the structure of the annexin XII hexamer described above, we propose a new model for the interaction of annexins with planar bilayers (Fig. 3b). This speculative model is not necessarily mutually exclusive with the current model (Fig. 3a). In fact, we speculate that both conformations of annexin XII would be present on bilayers. Our model postulates rearrangement of the local bilayer structure such that the charged phospholipid head groups interact favourably with the 18 type II  $\text{Ca}^{2+}$  sites situated on the perimeter of the annexin hexamer. This can be achieved by a local micelle-like structure of the phospholipid bilayer that would allow the annexin hexamer to span the planar bilayer while retaining hydrophilic interactions with the phospholipid head groups (Fig. 3b) very similar to the interactions proposed for the peripheral binding model (Fig. 3a). Such micelle-like bilayer structures have been proposed as the principal hydrophilic pores caused by electro-poration of lipid bilayers<sup>23</sup>.

A bilayer-spanning annexin XII hexamer is well suited to exhibit ion channel activity by virtue of its symmetrical funnel-shaped hydrophilic pore at the centre of the hexamer (see Fig. 1a and b). The most narrow part of the pore (about 8 Å diameter and 10 Å in length) is at the interface of the opposing trimers and is defined by six equivalent lysine residues (Lys 132). The region immediately adjacent to the mouth of the pore is defined by six copies of helix IIC and the following loop. The side chains of Lys 132 as well as Lys 128 and Lys 137 from each of the six annexin XII molecules provide this region with a basic environment, suggestive of a channel with a selectivity for anions. It is interesting to note that the pore region of annexin V in a similar hexamer contains glutamates 128, 129, 130 and aspartic acid 137 and thus would be acidic as expected for a cation-selective channel. The minimum diameter of this hypothetical annexin V pore (about 12 Å) is somewhat larger than the annexin XII pore because annexin V has a glycine at position 132.

In summary, we have determined the atomic structure of an annexin XII hexamer which is stabilized by unique intersubunit  $\text{Ca}^{2+}$  sites. Additional  $\text{Ca}^{2+}$  sites on the perimeter of the hexamer are proposed to mediate hydrophilic insertion of the protein complex into phospholipid bilayers. The structure of this bilayer-spanning complex provides a plausible mechanism by which annexins can form ion channels. The  $\text{Ca}^{2+}$  on the perimeter of the hexameric annexin also could bind to two different bilayers and thereby mediate annexin-catalysed vesicle aggregation<sup>3</sup> and fusion<sup>5</sup>. □

Received 23 June; accepted 2 October 1995.

1. Moss, S. E. (ed.) *The Annexins* (Portland, London, 1992).
2. Swairjo, M. A. & Seaton, B. A. *Rev. Biophys. biomolec. Struct.* **23**, 193–213 (1994).
3. Creutz, C. E. *Science* **258**, 924–931 (1992).
4. Pollard, H. B. & Rojas, R. E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2974–2978 (1988).
5. Rojas, E., Pollard, H. B., Haigler, H. T., Para, C. & Burns, A. L. *J. biol. Chem.* **265**, 21207–21215 (1990).
6. Berendes, R., Voges, D., Demange, P., Huber, R. & Burger, A. *Science* **262**, 427–430 (1993).
7. Demange, P. et al. *Trends biochem. Sci.* **19**, 272–276 (1994).
8. Schlaepfer, D. D. et al. *J. biol. Chem.* **267**, 9529–9539 (1992).
9. Voges, D. et al. *J. molec. Biol.* **238**, 199–213 (1994).
10. Huber, R. et al. *J. molec. Biol.* **223**, 683–704 (1992).
11. Weng, X. et al. *Protein Sci.* **2**, 448–458 (1993).
12. Creutz, C. E., Pazoles, C. J. & Pollard, H. B. *J. biol. Chem.* **254**, 553–558 (1979).
13. Jost, M., Weber, K. & Gerke, V. *Biochem. J.* **298**, 553–559 (1994).
14. Holland, D. R. et al. *J. biol. Chem.* **265**, 17649–17656 (1990).
15. Ahn, N. G. et al. *J. biol. Chem.* **262**, 18657–18663 (1988).
16. Zaks, W. J. & Creutz, C. E. *Biochemistry* **30**, 9607–9615 (1991).
17. Concha, N. O., Head, J. F., Kaetzel, M. A., Dedman, J. R. & Seaton, B. A. *FEBS Lett.* **314**, 159–162 (1992).
18. Neumann, J.-M., Sanson, A. & Lewit-Bentley, A. *Eur. J. Biochem.* **225**, 819–825 (1994).
19. Creutz, C. E., Pazoles, C. J. & Pollard, H. B. *J. biol. Chem.* **253**, 2858–2866 (1978).

20. Mosser, G. *J. molec. Biol.* **217**, 241–246 (1991).
21. Huber, R., Römisch, J. & Paques, E.-P. *EMBO J.* **9**, 3867–3874 (1990).
22. Burger, A. et al. *J. molec. Biol.* **237**, 479–499 (1994).
23. Weaver, J. C. *Ann. N.Y. Acad. Sci.* **720**, 141–152 (1993).
24. Hille, B. *Ionic Channels of Excitable Membranes* 2nd edn (Sinauer, Sunderland, MA, 1992).
25. Smart, O. S., Wallace, B. A. & Goodfellow, J. M. *Biochem. Soc. Trans.* **22**, 146S (1994).
26. Leckband, D. E., Israelachvili, J. N., Schmitt, F.-J. & Knoll, W. *Science* **255**, 1419–1421 (1992).
27. Leslie, A. G. W. in *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography* No. 26 (Daresbury Laboratory, Warrington, 1992).
28. Tong, L. & Rossmann, M. G. *Acta crystallogr.* **A46**, 783–792 (1990).
29. Sack, J. S. *J. molec. Graph.* **6**, 224–225 (1988).
30. Brünger, A. T. *X-PLOR, Version 3.1, A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, 1992).

ACKNOWLEDGEMENTS. We thank our colleagues for critical reading and suggestions on the manuscript, and M. Levitt for the use of his molecular modelling workstations. SSRL is funded by the Department of Energy, Office of Basic Energy Sciences. The Biotechnology Program is supported by the NIH, Biomedical Research Technology Program, Division of Research Resources. Further support is provided by the Department of Energy, Office of Health and Environmental Research. This work was also supported by grants (to H.T.H.) from the American Cancer Society and National Science Foundation. W.S.M. was supported by a predoctoral training grant.



# Mapping the genome one molecule at a time — optical mapping

Akhtar H. Samad, Wei Wen Cai, Xinghua Hu, Benjamin Irvin, Junping Jing, Jason Reed, Xun Meng, John Huang, Edward Huff, Brett Porter, Alex Shenkar, Thomas Anantharaman, Bhubaneswar Mishra, Virginia Clarke, Eileen Dimalanta, Joanne Edington, Catharina Hiort, Roderic Rabbah, John Skiada & David C. Schwartz

DESPITE rapid progress in the human genome project effort, there is little doubt that new conceptual approaches are needed before whole genome-based analyses become routine. This is largely true because current molecular biological approaches were developed primarily for characterization of single genes, not entire genomes, and, as such, are not ideally suited to population-based molecular genetics.

Attempts at identifying and characterizing various genetic loci have been facilitated enormously by physical mapping studies of the human genome, involving approaches such as restriction endonuclease and hybridization analyses. Restriction maps help determine the spatial relationships of specific loci, by providing precise genomic distances and the order of physical landmarks, and contribute to relatively unambiguous clone characterization that is useful in contig formation and in establishment of minimal tiling paths for sequencing efforts. In comparison, tedious hybridization-based approaches generally yield low complexity maps, and ordered sequence-based landmarks such as Sequence Tagged Sites (STSs) lack the informativeness of high-resolution restriction maps, despite the simplicity and speed of the latter approach. Perhaps because they still utilize electrophoresis, the associated techniques for the generation of restriction maps have changed little over the past ten years, and fully automated restriction mapping systems have yet to be widely adopted. To help overcome these shortcomings, our laboratory developed the first practical non-electrophoretic genomic physical mapping approach, optical mapping<sup>1-4</sup>.

Optical mapping is a single molecule methodology for the rapid production of ordered restriction maps from single DNA molecules. Ordered restriction maps were constructed originally from yeast chromosomes<sup>1</sup> by using fluorescence microscopy to visualize restriction endonuclease cutting events on single, fluorochrome-stained DNA molecules. Restriction enzyme cleavage sites were visible as gaps that appeared as the DNA fragments relaxed. Relative fluorescence intensity or appar-

ent length measurements of the restriction fragments were then used to construct the final restriction map. In the original method<sup>1</sup>, fluorescently labelled DNA molecules were elongated in a flow of molten agarose containing restriction endonucleases, generated between a coverslip and microscope slide, and the resulting cleavage events were recorded by fluorescence microscopy as time-lapse digitized images. Nevertheless, improvements in sizing sensitivity and throughput were required if a wide range of cloning vectors (cosmid, bacteriophage, yeast artificial chromosomes (YAC)) were to be analysed by optical mapping. Given the utility of such clones in the construction of contigs for mapping and sequencing projects, we developed a second generation optical mapping approach, which dispensed with

ments have lowered the limit of resolution to about 300 bp. The reported relative sizing accuracy of 5 per cent for fragments >5 kb was comparable to sizing accuracy by agarose gel electrophoresis. Sizing accuracy is a function of the number of molecules analysed, and accuracy to within a few base pairs on kilobase-range restriction digest fragments of lambda bacteriophage has been achieved, representing higher accuracy than is achievable by agarose gel electrophoresis.

A large fraction of the human genome is presently covered by YAC contigs<sup>5</sup>, a notable first step towards generation of the first genome-wide physical map. However, due to the high frequency of rearrangements or chimerism among YACs, and the low complexity of fingerprints generated by hybridization approaches, extensive high-resolution restriction maps of YACs have not been widely generated. Ordered restriction maps of YACs have been generated in this laboratory by optical mapping<sup>4</sup>, with overall relative sizing accuracies that are comparable to routine pulsed-field gel electrophoretic analysis (see fig.). The approach taken by this laboratory involved combining the fluid turbulence damping properties of molten agarose, with the enzymatic accessibility of surface mounting, by fixing YACs in molten agarose onto derivatized glass surfaces<sup>4</sup>.

Using optical mapping, this laboratory has generated ordered restriction maps for the Beckwith-Wiedeman locus in humans (in collaboration with Housman's group at the Massachusetts Institute of Technology), the *Brca2* locus (in collaboration with Fisher's group at Columbia University), and the mouse olfactory locus (in collaboration with Axel's group at Columbia University). Optical Maps are currently being generated from phage, cosmid, YAC and bacterial artificial chromosome (BAC) clones, utilizing approaches that obviate the glass sandwich layer described above, and which can accommodate the high-throughput requirements of whole-genome analysis. Before the development of optical mapping, restriction maps were made by

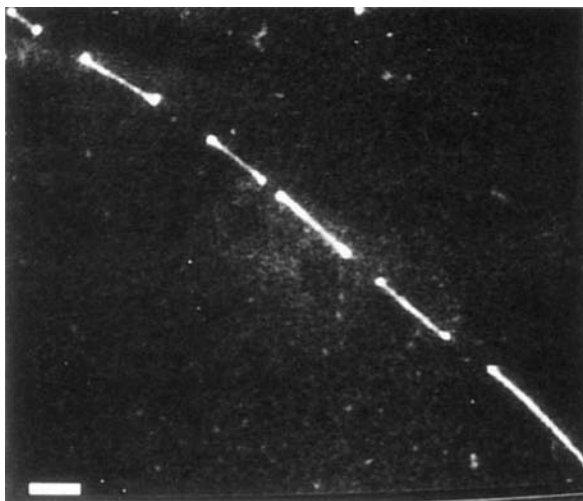


Image of restriction enzyme digested YAC clone: YAC clone 6H3, derived from human chromosome 11, digested with the restriction endonucleases *Eag* I and *Mlu* I, stained with a fluorochrome, and imaged by fluorescence microscopy (see ref. 4 for details). Six fragments are generated from the 425-kb parent molecule. The bar represents 10  $\mu$ m.

agarose and time-lapse imaging, and involved fixing elongated DNA molecules onto polylysine-treated glass surfaces. For analysis of lambda clones, DNA samples were fixed onto derivatized glass surfaces by sandwiching between a treated coverslip and glass slide, and a charge coupled device (CCD) camera was used to image molecules from 28 kilobases (kb) down to 800 base pairs (bp)<sup>3</sup>; more recent experi-

tedious gel electrophoresis-based methods. Although maps have been made successfully for the genomes of *Escherichia coli*<sup>5</sup>, *Saccharomyces cerevisiae*<sup>6</sup>, and *Cenorhabditis elegans*<sup>7</sup>, numerous limitations of this approach remain, foremost among them being the need for higher throughput that is less labour intensive. Automation of optical mapping also holds enormous promise for miniaturization, with expected increases in throughput and reductions in cost. We expect that the

advantages of this technique will facilitate completion of the initial objectives of the human genome project, and on a population scale both mapping of disease genes and detection/mapping of genetic alterations afflicting the human genome. □

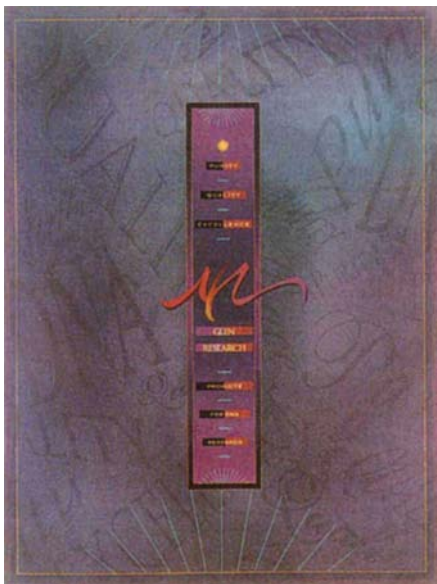
A.H.S. and D.C.S. are with the Dept. of Pathology, Cornell Medical College, New York, New York, 10021 USA and Dept. of Chemistry, New York University, New York, New York 10003, USA, respectively. For more information fill in reader service number 100.

1. Schwartz, D. C. et al. *Science* **262**, 110–114 (1993).
2. Wang Y et al. *Proc. natn. Acad. Sci. U.S.A.* **92**, 165–169 (1995).
3. Meng, X., Benson, K., Chada, K., Huff, E. & Schwartz, D.C. *Nature Genet.* **9**, 432–438 (1995).
4. Cai, W., Housman, D.E., Wang, Y. & Schwartz, D.C. *Proc. natn. Acad. Sci. U.S.A.* **92**, 5164–5169 (1995).
5. Kohara, Y., Akiyama, K. & Isono, K. *Cell* **50**, 495–508 (1987).
6. Olson, M. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7821–7825 (1986).
7. Coulson, A., Sulston, J., Brenner, S. & Karn, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7826–7830 (1986).
8. Cohen, D., Chumakov, I. & Weissenbach, J. *Nature* **366**, 698–701 (1993).

# Synthesis and sequencing

**A sequence-analysis system, software packages, DNA purification supplies, Internet information and a custom oligonucleotide synthesis service are some of the items covered in this product review feature.**

THE new Glen Research catalogue now available from Cambio contains over 40 pages of products for DNA and RNA research (*Reader Service No. 101*). The catalogue contains information on Sterling Synthesis Products, Minor Base Phosphoramidites, products for DNA modification



**1995–96 Glen Research catalogue.**

and labelling, RNA and 2'-OMe RNA monomers. Products are offered in a choice of vials or columns, suitable for all synthesizers and are delivered with a certificate of analysis.

Genosys has just introduced its new GeneScout kit for rapid recombinant screening and template preparation for manual and automated DNA sequencing (*Reader Service No. 102*). The kit is said to allow rapid analysis of recombinant colonies without doing plasmid minipreps. Bacterial colonies or viral-plaque lysates can be used directly in PCR amplification. The manufacturer states that the sequencing template generated is single-stranded

for optimal use with sequencing polymerases. Sequencing templates produced using the kit are said to be ideal for both manual and automated sequencing techniques. It makes use of AffiniTip streptavidin microcolumns to capture biotinylated PCR products for strand separation. Several kits are available with different sets of PCR primers for use with a variety of common vectors, including lacZ-based vectors,  $\lambda$ gt10,  $\lambda$ gt11 and baculovirus cloning vectors.

## Synthesizers and sequencers

The LI-COR infrared DNA sequencing and analysis systems are designed to increase the accuracy and affordability of DNA analysis in the laboratory (*Reader Service No. 103*). The systems combine infrared fluorescence chemistry with the latest laser technology and software to produce high-performance systems for DNA sequencing and genetic analysis. Infrared DNA analysis is said to offer consistent, maintenance-free operation with a reliable, solid-state laser and detector. There are three models available to suit a variety of user needs. The systems include Base ImageIR 2.0 software featuring Quick SequencIR, a new software tool that automates electrophoresis and base calling. A single click performs the entire sequencing process: electrophoresis image data collection, lane finding and base calling.

Oncor's new Fidelity DNA sequencing system provides manual sequencing of single-stranded (ss) and double-stranded (ds) DNA templates using a proprietary T4 polymerase system (*Reader Service No. 104*). The system is designed to provide long, resolvable reads on dsDNA with even band intensities and fewer pause sites. It is configured into a streamlined twelve-component kit, which requires no predilution of the polymerase or reaction buffers, and is packaged in a reusable

benchtop cooler. Additionally, a quick-denature dsDNA protocol is provided to shorten sample preparation time. The system uses a 3'-5' exonuclease-deficient form of T4 DNA polymerase in conjunction with T4 ssDNA-binding protein and T4 polymerase accessory factors to provide exceptional manual sequencing of ss and ds templates.

PerSeptive Biosystems' Expedite nucleic acid synthesizer has two independent columns, allowing operation of different scales and different protocols simultaneously (*Reader Service No. 105*). The microfluidics plate technology is designed to achieve low-reagent usage — total consumption per cycle is <4.5 ml per cycle at the 0.05-, 0.2- and 1.0- $\mu$ mol scales. It has a total run time of 3.5 minutes, no matter how many columns are running. This Windows-based workstation offers an



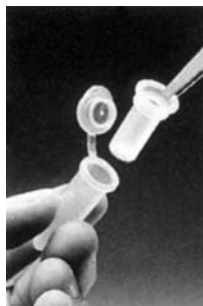
## Expedite from PerSeptive Biosystems.

internal database, expanded editing and the multiple oligonucleotide synthesis system (MOSS). This system is said to provide unattended synthesis of up to 16 oligonucleotides, allowing overnight operation and increased productivity. Independent columns allow control of individual protocols, even while the system is operating. With a modular design that attaches to the top of the system, MOSS requires no extra bench space.

## PRODUCT REVIEW

### Separation and purification

Amicon's new Micropure-EZ is said to provide an easy, phenol-free means of removing restriction- and DNA-modifying enzymes from up to 250  $\mu$ l of reaction mixtures containing double-stranded (ds) DNA (*Reader Service No. 106*). The membrane in this **centrifugal device** has a high affinity for protein but not for dsDNA. In a single 30-s spin, enzymes are selectively adsorbed. DNA is recovered in a vial — undiluted and enzyme-free — and is said to be typically 90 per cent pure or higher. The company claims that it is not affected by DNA size or the binding capacity of the adsorption matrix. It is microcentrifuge compatible and can be spun at 14,000g.



Micropure-EZ.

Amresco now offers **gel matrices for polyacrylamide gel electrophoresis** that are designed to provide higher resolution of nucleic acids (*Reader Service No. 107*).

PAGE-Plus and Gene-PAGE Plus are intended to offer researchers a versatile, high-resolution gel that is more resistant to mechanical manipulations than standard acrylamide-bis-acrylamide mixtures. These mixtures are intended for applications that require additional sequence information and resolution of low molecular weight nucleic acids. These solutions are provided in a ready-to-use format, eliminating the hazards associated with handling powdered acrylamide derivatives.

### Sequencing software

Gene Codes has announced that Hitachi Software Engineering will sell its Sequencer **sequencing software** in Japan (*Reader Service No. 108*). The software copies strings of bases and assembles matching fragments to identify contigs. Sequencer is a pattern matching program that is designed to expedite assembly of hundreds or even thousands of these fragments, a process that formerly took hours and now is said to take only seconds to perform. Hitachi's product line includes: DNA sequence-analysis software (for Macintosh and PC computers), high-speed homology search boards, electromagnetic digitizers and the FMBIO-100, a filmless non-radioactive imaging device.

IntelliGenetics, the bioinformatics division of the Oxford Molecular Group, has incorporated the popular FASTA searching algorithm into its release of PC/GENE 6.85 **sequence analysis software** (*Reader Service No. 109*). This bioinformatics software has extensive protein analysis capabilities for use by researchers working on desktop personal computers, and should prove useful for understanding the biological significance of DNA and protein sequences. Users can now conduct sequence similarity searches on EMBL, Unique GenBank, SWISS-PROT and personal databases using the FASTA searching algorithm. Developed to increase control and sensitivity of searches, users can choose between PAM, BLOSUM and Genetic Code similarity matrices, and set k-tuple and gap penalties. PC/GENE now displays up to 2,000 scores and alignments, along with a complete histogram of Init1 and InitN scores for the entire database searched.

Bio Image has announced the availability of **DNA sequence film reader software and sequence assembly manager software**, used for electrophoresis image analysis and assembling DNA fragments into contiguous sequences (*Reader Service*

ADVERTISEMENT

### DNA/RNA

#### Isolation and Purification

- **Tri Reagent™**— Isolates DNA, RNA and protein in a single step.
- **PlasmidPURE™** DNA Miniprep Kit— Inexpensively isolates, high yields of pure plasmid DNA in less than 20 minutes.
- **NucleiClean™** Nucleic Acid Rapid Isolation Kit— Recovers up to 100% of DNA from solution and 80% from agarose gels while eliminating the need for phenol extractions and ethanol precipitations.



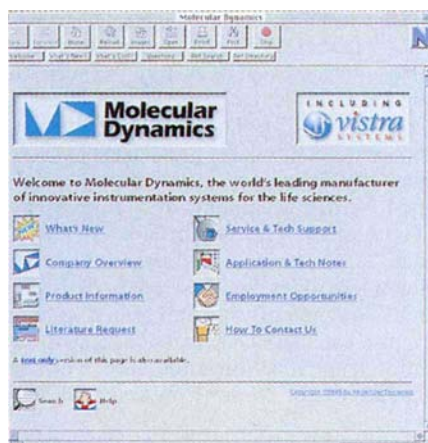
**"Be 3 times more productive in isolation. Sigma BioSciences offers 3 new products for easy nucleic isolation and purification."**



No. 110). DNA sequence film reader software automatically locates lanes and bands, corrects for sequencing anomalies, such as band smearing, and reports the sequence. Onscreen viewing can be improved by zooming in on difficult areas, selecting higher contrast colours or using one of three quantitative decision support tools. The sequence assembly manager automatically assembles fragments into contiguous sequences. It is capable of importing and assembling as many as 1,000 fragments into a single contiguous sequence up to 50 kb in length. It is designed to eliminate time-consuming manual verification of ambiguous regions. The software automatically retrieves ambiguous portions of the original image in seconds, and displays them onscreen in the correct forward, reverse or complementary order.

### Internet news

Molecular Dynamics has launched a world wide web site on the Internet at location <http://www.mdyn.com> (Reader Service No. 111). This site details a complete range of life science products including fluorescent imaging, filmless autoradiography, densitometry, confocal microscopy, and DNA sequencing and



The Molecular Dynamics 'home page'.

automation. Additional site pages include application and technical notes, scientific meetings and new product press releases.

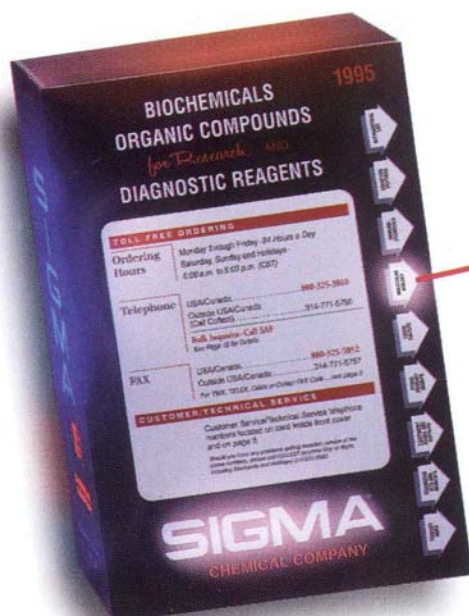
The Institute of Genomic Research (TIGR) and the American Type Culture Collection (ATCC) have announced the creation of the TIGR/ATCC human cDNA special collection (Reader Service No. 112). This will be a repository of human cDNA materials, representing the majority of human genes. Human cDNA

clones from TIGR will be made available to the scientific community through the distribution systems provided by ATCC. Consistent with current ATCC policy, materials will be available to the scientific community without restriction (including all human cDNA clones previously available from ATCC), clones restricted to not-for-profit institutions and proprietary clones available to users of the Human cDNA Database (HCD) at TIGR or researchers who sign a material transfer agreement. Information about human cDNA clones will be provided through the Internet. <http://www.atcc.org>

### Gel documentation

Ultra Violet Products has introduced a new series of special filters for its entire range of gel documentation systems (GDS) (Reader Service No. 113). The filters allow gel documentation users to employ SYBR green gel stains instead of traditional ethidium bromide stains — SYBR stains are not carcinogenic and offer enhanced DNA and RNA detection limits in agarose or polyacrylamide gels. UVP's series of gel documentation and analysis systems, the GDS 2000, 5000, 7500 and 8000, usually require an appropriate transilluminator to irradiate the gels

### ADVERTISEMENT



"Put yourself in splendid isolation. Order here."

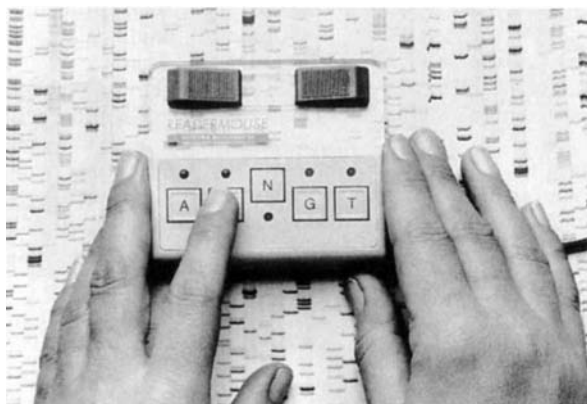
At Sigma BioSciences, we've just tripled our reputation as your #1 source for DNA and RNA isolation and purification products with three new products that will deliver quicker and easier isolation of nucleic acids: Tri Reagent™, NucleiClean™, Nucleic Acid Rapid Isolation Kit, PlasmidPURE™, DNA Miniprep Kit.

So put yourself in splendid isolation! Like all Sigma BioSciences products, Tri Reagent, NucleiClean and PlasmidPURE can easily be ordered by calling 1-800-325-3010 (USA/Canada) or 314-771-5750 (call collect outside USA and Canada).

**SIGMA**  
BIO SCIENCES™

under investigation with ultraviolet (UV) light. Depending on the analysis, various transilluminator sizes operating at a selection of wavelengths (365 nm, 302 nm and 254 nm) are available. The company offers single-, double- and triple-UV, white-light/UV and white-light units. The manufacturer states that at 254 nm, as little as 20 pg of dsDNA can be detected with SYBR Green I. SYBR I Green is used for DNA detection and SYBR II Green is used for RNA detection and non-isotopic DNA single-strand, conformation polymorphism analysis (SSCP).

Readermouse from Integra Biosciences is designed for **direct computer entry and manual editing of nucleic acid sequences** (Reader Service No. 114). DNA sequences can now be read directly from an X-ray film into a computer without the need for a keyboard. The new device can commu-



**Readermouse communicates directly with a computer.**

nicate directly with a computer. This version is suitable for use with Windows and Macintosh systems and enables DNA sequences to be edited, checked, processed, stored and exported.

The **agarose gel DNA extraction kit**, recently introduced by Boehringer Mannheim, is designed to overcome the problem of purifying sufficient quantities of DNA (particularly large-fragment DNA), a problem for traditional gel-purification methods (Reader Service No. 115). The new kit is designed to optimize the recovery of DNA fragments from standard and low melting point agarose gels. Efficient isolation is found with fragment sizes ranging from 400 bp to 5 kb. Bound molecules are separated and purified in simple washing steps, and the nucleic acid is removed by elution from the binding matrix using a low-salt buffer or water. The fragments isolated using this kit can then be ligated for cloning purposes or used to generate probes in random-prime labelling or nick translation reactions. The manufacturer states that high purity and an 80 per cent yield can be expected for all sizes of fragments.

## DNA services

Life Technologies has announced the release of Gibco BRL custom primers, an **oligonucleotide synthesis service** (Reader Service No. 116). These custom primers are manufactured by the Parallel Array Synthesis technology, a new process developed by the company and ProtoGene Laboratories. The system is said to produce large arrays of different custom oligonucleotides simultaneously. This system is designed to use integrated quality control, extensive washes, collection of the entire trityl reaction and direct vertical photometry for accurate process analysis.

Larova now offers a **custom DNA sequencing service** of M13 DNA, plasmids and PCR products (Reader Service No. 117). The company utilizes the LICOR 4000 DNA sequencer, which is said to have high sensitivity and accuracy. Two services may be selected: 'lab quality', which is appropriate for getting first information about subclones (identity, insert orientation) or PCR products and 'publication quality', which comprises at least one sequence reading on either strand for achieving the best accuracy. Documentation is included.

## Miscellaneous

Epicentre now offers its 1995/1996 **molecular and cellular biology catalogue**, featuring new enzymes, reagents and kits (Reader Service No. 118). The catalogue describes products and procedures for isolation and purification of DNA and RNA, nucleic acid sequencing, *in vitro* transcription, molecular cloning and analysis, and protein studies, as well as a wide range of thermostable enzymes. The Epicentre Forum, a free bimonthly newsletter on new tools and techniques for molecular biology, is also available.

Typically, **stabilized <sup>35</sup>S nucleotides** have not been suitable for use in certain applications, including PCR and end-labelling techniques. NEN Life Science Products claims to have solved the problem with Easytide <sup>35</sup>S dATP. (Reader Service No. 119). With a new patented proprietary buffer, Easytide <sup>35</sup>S dATP is said to be compatible with all applications. The radiochemical stabilizer slows Easytide <sup>35</sup>S dATP decomposition, allowing storage at 4 °C, which does not interfere with PCR and DNA/RNA-labelling techniques. The new <sup>35</sup>S nucleotide is shipped as a ready-to-use liquid and is colour-coded sapphire blue to make it easy to see, for safer and more accurate pipetting. □

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

## ADVERTISEMENTS

### FLUORESCENT PROBES

DNA/RNA Stains *Fura-2 AM*  
Ion Probes *1 mg DM 141.10*  
pH Indicators  
Enzyme Substrates *BCECF-AM*  
Fluorescent Labels *1 mg DM 78.00*  
Membrane Probes

EXCELLENT QUALITY - COMPETITIVE PRICES

LAMBDA FLUORESCENZTECHNOLOGIE GmbH  
Grottenhofstr. 3, A-8053 Graz, AUSTRIA.  
Tel.: ++43316/265352 Fax: ++43316/278586  
In USA and CANADA call: 814-359-5064

Reader Service No.3

### OLIGONUCLEOTIDES

**£1 per base**

- stringent quality control
- express turnaround
- column purified: ready to use

**By research labs for research labs**

**Molecular Biology Services**

King's College School of Medicine and Dentistry  
Rayne Institute, 123 Coldharbour Lane, London SE5 9NU  
Tel. 0171-346 3126 Fax: 0171-733 3877

Reader Service No.10

### DNA SEQUENCING

**£45 per reaction**

- PCR products
- single strand DNA
- M13 inserts
- double strand DNA
- plasmid inserts
- no result? no charge!

**By research labs for research labs**

**Molecular Biology Services**

King's College School of Medicine and Dentistry  
Rayne Institute, 123 Coldharbour Lane, London SE5 9NU  
Tel: 0171-346 3126 Fax: 0171-733 3877

Reader Service No. 11

### G418

**Cell Culture Tested**

**1 gram \$69**

**5 grams \$250**

**Quotes on request**

**Mediatech Inc**

13884 Park Center Road, Herndon,  
Va. 22071, USA

**800 - Cellgro**

Phone: 703-471-5955

Fax: 703-471-0363

Reader Service No.14